

Évolution des moyennes trimestrielles en 9^e du cycle d'orientation : quels en sont les déterminants ?

*Étude réalisée en collaboration avec
le groupe Édumétrie de la SSRE*



**Franck Petrucci, Anne Soussi,
François Rastoldo, Edith Guilley,
Christian Nidegger**

Novembre 2015

Évolution des moyennes trimestrielles en 9^e du cycle d'orientation : quels en sont les déterminants ?

*Étude réalisée en collaboration avec
le groupe Édumétrie de la SSRE*

**Franck Petrucci, Anne Soussi,
François Rastoldo, Edith Guilley,
Christian Nidegger**

Novembre 2015

Fin des travaux : Septembre 2015

Remerciements

Nos remerciements s'adressent, en premier lieu, à nos collègues du groupe Édumétrie avec qui la collaboration s'est révélée extrêmement enrichissante et stimulante. Merci plus particulièrement à Daniel Bain qui nous a fait profiter de ses connaissances, de son expertise et de son regard critique. Merci aussi à Gianreto Pini pour ses précieux conseils méthodologiques ainsi qu'à Emiel Reith et Dagmar Hexel pour leur contribution importante à la réalisation de chacune des phases de ce projet.

Enfin, nous tenons également à remercier la direction générale de l'enseignement obligatoire (DGEO) d'avoir mis à notre disposition les données nécessaires pour la réalisation de cette recherche.

Compléments d'information :

Franck Petrucci
Tél. +41/0 22 546 71 22
franck.petrucci@etat.ge.ch

Anne Soussi
Tél. +41/0 22 546 71 39
anne.soussi@etat.ge.ch

François Rastoldo
Tél. +41/0 22 546 71 36
francois.rastoldo@etat.ge.ch

Edith Guilley
Tél. +41/0 22 546 71 51
edith.guilley@etat.ge.ch

Christian Nidegger
Tél. +41/0 22 546 71 19
christian.nidegger@etat.ge.ch

Responsable de l'édition :

Narain Jagasia
Tél. +41/0 22 546 71 14
narain.jagasia@etat.ge.ch

Internet :

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

Diffusion :

Service de la recherche en éducation (SRED)
12, quai du Rhône - 1205 Genève
Tél. +41/0 22 546 71 00
Fax +41/0 22 546 71 02

Document 15.017

*Le contenu de ce document n'engage que la responsabilité
du Service de la recherche en éducation.*

Table des matières

En préambule.....	5
L'essentiel en bref.....	7
1. Introduction	9
Le système scolaire	9
Les enseignants	10
Les élèves.....	11
2. Objectifs de la recherche et méthodologie.....	13
3. Organisation du rapport.....	15
4. Échantillonnage des données.....	17
La volée 2012.....	17
La volée 2007.....	17
Différences entre les volées 2007 et 2012	18
Le choix des disciplines	18
5. Présentation du modèle multiniveau de croissance	19
6. Présentation des résultats	21
Le cas des moyennes trimestrielles de mathématiques en 9 ^e CO pour la volée 2012.....	21
Le cas des moyennes trimestrielles de français et d'histoire en 9 ^e CO pour la volée 2012	32
Évolution simultanée des moyennes trimestrielles de français, mathématiques et histoire en 9 ^e CO pour la volée 2012.....	35
Une autre organisation du CO : le cas des moyennes trimestrielles en 7 ^e CO pour la volée 2007 ...	37
7. Synthèse, discussion et perspectives.....	45
Distributions différenciées entre R1, R2 et R3 (mathématiques).....	45
Mêmes types d'évolution d'un trimestre à l'autre dans les trois regroupements	46
Effet du genre et de l'origine sociale	46
Effet « positif » d'être en R1 ou R2 à niveau initial comparable ?	47
Effet du temps	47
Des résultats similaires en mathématiques, français et histoire	47
Comparaison des deux systèmes (volées 2007 et 2012)	48
Références bibliographiques	49

En préambule

La présente recherche a été initiée par le groupe Édumétrie de la Société suisse pour la recherche en éducation (SSRE), sur mandat de la direction générale de l'enseignement obligatoire (DGEO) et a donné lieu à une collaboration avec le service de la recherche en éducation (SRED) du canton de Genève. La finalité de ces travaux était double :

- ♦ étudier l'intérêt du modèle multiniveau de croissance, se familiariser avec la méthodologie et la mettre à l'épreuve du terrain. Relativement récent, ce type d'outil statistique est en effet encore peu connu des chercheurs et des autres acteurs du champ de l'éducation. Il est important d'en saisir les possibilités et les limites pour l'utiliser de façon adéquate ou pour comprendre et interpréter correctement les résultats qui en sont issus et discuter les conclusions avec les responsables scolaires ;
- ♦ contribuer à l'analyse du fonctionnement du CO dans un domaine crucial pour l'orientation.

A la suite de ces travaux, le SRED et le groupe Édumétrie de la SSRE ont organisé conjointement, le 20 mars 2015 au SRED, une journée d'étude sur le thème « Analyser l'évolution des résultats des élèves : introduction au modèle multiniveau de croissance ». Lors de cette journée, financée par l'Association suisse des sciences humaines (ASSH), la présente recherche a donné lieu à une communication dont le contenu est disponible, avec l'ensemble des autres travaux exposés, à l'adresse suivante : <https://www.geneve.ch/recherche-education/>.

L'essentiel en bref

Cette étude, réalisée en collaboration avec le groupe ÉduMétrie de la SSRE, a pour thème l'évolution des moyennes trimestrielles au début du cycle d'orientation (CO). Évaluer les acquis des élèves et en rendre compte par l'attribution de notes est toutefois un phénomène complexe qui peut être sujet à un certain nombre de biais liés au contexte ou à la classe, à l'enseignant et à ses valeurs ou normes, ou encore aux caractéristiques individuelles des élèves. En effet, la notation ne résulte pas d'une simple application d'un barème de correction.

Cette recherche, centrée sur l'évolution des moyennes trimestrielles dans trois disciplines (mathématiques, français et histoire), porte sur les résultats au début du CO, en 2012 (c'est-à-dire une année après la mise en place de la nouvelle organisation du CO) ainsi qu'en 2007 à titre comparatif. Elle a pour objectif de répondre à différentes questions : quel type d'évolution les moyennes suivent-elles tout au long de l'année (progression, régression ou stabilité) ? Observe-t-on des différences selon les élèves (profils) en fonction de leur niveau initial ou de différentes caractéristiques (genre, première langue parlée ou origine socio-économique) ? Constate-t-on des variations en fonction de la discipline évaluée et selon le rôle qu'elle joue dans la sélection ou l'orientation ? Les résultats varient-ils en fonction de l'organisation du CO en 2012 (avec trois regroupements différenciés) et en 2007 (avec deux regroupements¹ et des classes hétérogènes) ? Les changements d'organisation et de seuil de suffisance ont-ils un impact sur les moyennes trimestrielles et leur évolution tout au long de l'année ?

Afin de mesurer l'évolution des moyennes trimestrielles au cours de l'année, nous avons retenu une approche statistique particulièrement adéquate pour observer ce genre de phénomène : le modèle multiniveau de croissance.

Les résultats montrent qu'en 2012, l'évolution des moyennes trimestrielles est pratiquement identique pour les trois disciplines. On peut donc choisir d'illustrer d'abord ce résultat par l'exemple des mathématiques. Les moyennes ont tendance à baisser entre le premier et le troisième trimestre, quelle que soit la filière considérée, comme si les enseignants au début de l'année se montraient plus indulgents et ajustaient ensuite leur niveau d'exigence en fin d'année au moment de la certification et de l'orientation. Une autre explication peut être l'intégration des résultats aux évaluations externes dans la moyenne du troisième trimestre, résultats le plus souvent inférieurs aux résultats des évaluations internes. Soulignons toutefois que les variations au cours de l'année sont relativement faibles. Par ailleurs, on observe des moyennes plus élevées dans le regroupement 3 (R3) aux exigences les plus élevées que dans les deux autres regroupements (R1 et R2). Cette différence d'utilisation de l'échelle des notes peut sans doute s'expliquer par un effet des dispositions réglementaires relatives aux réorientations (des moyennes supérieures à 4.8 permettent une réorientation vers une section plus exigeante). Il est possible qu'une notation différente soit pratiquée entre les regroupements, en particulier des notes moins élevées dans les regroupements aux exigences les plus faibles afin de ne pas donner aux élèves de faux espoirs en vue d'une éventuelle promotion². Une autre explication peut être liée aux représentations des exigences concernant les objectifs de 9^e de la part des enseignants qui prendraient comme référence les exigences du regroupement 3.

L'analyse multiniveau de croissance montre que la majeure partie de la variance totale des résultats en mathématiques (plus de 93%) est constituée par des différences interindividuelles et que seule une petite part (près de 7%) correspond à des variations de moyennes enregistrées d'un trimestre à l'autre pour un même élève. Au niveau des caractéristiques interindividuelles de type sociodémographiques, seul le milieu d'origine socioéconomique des élèves a un impact sur les résultats « toutes choses

¹ Il existait un troisième regroupement (regroupement C), scolarisant 3% des élèves, uniquement dans certains établissements.

² Il faut souligner que les élèves ayant fait l'objet d'une réorientation ne font pas partie de l'échantillon analysé étant donné que l'on ne dispose pas de l'ensemble de leurs résultats trimestriels.

égales par ailleurs » (c'est-à-dire lorsqu'on contrôle les autres variables) : il est positif quand les élèves proviennent d'un milieu favorisé, et négatif quand ils proviennent d'un milieu modeste ou défavorisé. On n'observe pas d'effet du genre ni de la première langue parlée sur les moyennes trimestrielles en mathématiques.

Les caractéristiques scolaires, à savoir le niveau initial en mathématiques (résultats en fin de primaire, 8P) ainsi que le résultat au test de raisonnement non verbal (préorientation), ont un effet positif sur les résultats trimestriels : plus un élève a un niveau élevé en mathématiques ou au test de raisonnement non verbal à la fin de l'école primaire, meilleures seront ses moyennes trimestrielles en 9^e CO (effet cumulatif des connaissances). Un autre élément a un effet positif sur les résultats : l'appartenance au regroupement R1 ou R2 par rapport au regroupement R3, c'est-à-dire que « toutes choses égales par ailleurs » (notamment à même niveau au primaire), un élève obtient une moyenne trimestrielle de mathématiques plus élevée lorsqu'il est scolarisé au CO en R1 ou R2.

On observe également des effets d'interaction entre d'une part, le temps et le niveau scolaire de 8P et d'autre part, le temps et l'appartenance à R1 et R2. Ainsi, si dans tous les regroupements, les moyennes ont tendance globalement à baisser au cours de l'année scolaire, cette baisse est toutefois moins élevée, toutes choses égales par ailleurs, lorsqu'un élève est scolarisé en R1 ou en R2 et qu'il possède un niveau initial de mathématiques plus élevé.

Un autre résultat intéressant concerne les profils de notes dans les trois disciplines : les corrélations entre moyennes sont plus élevées au troisième trimestre qu'au premier. Les enseignants connaissent de mieux en mieux le niveau de leurs élèves au cours de l'année scolaire, ce qui permet probablement un affinement de la notation. De plus, les positions respectives des élèves au sein de la classe ont tendance à se stabiliser. Il y aurait comme un effet de normalisation.

Comme annoncé plus haut, on constate les mêmes résultats dans les analyses réalisées sur les moyennes des deux autres disciplines : le français, qui joue également un rôle important dans l'orientation et la sélection, et l'histoire, dont la dotation horaire est moindre et le rôle moins prépondérant dans l'orientation. Ainsi, les moyennes ont tendance à légèrement baisser au cours de l'année. Les résultats trimestriels sont également plus élevés en R3 qu'en R2 et R1.

Les parts de variances respectives du niveau 1 (mesures) et du niveau 2 (élèves) se répartissent de la même manière qu'en mathématiques, la part majeure étant celle de niveau 2. Comme pour les mathématiques, l'analyse multiniveau de croissance met en évidence un effet du niveau initial et de l'origine socioéconomique. Par contre, en français, le genre joue également un rôle au profit des filles et en histoire, la première langue parlée au profit des élèves francophones. On constate également des effets d'interaction entre le temps et le regroupement d'une part, et entre le temps et le genre d'autre part.

Ces résultats similaires entre les trois disciplines, dont deux sont des disciplines principales, traduisent probablement l'existence de profils de compétences transversales, d'aptitude générale.

Les analyses réalisées sur les données de 2007 donnent lieu aux mêmes types de résultats qu'en 2012, à l'exception près que les moyennes sont systématiquement plus élevées en 2012. Ces résultats s'expliquent certainement par l'augmentation du seuil de suffisance cette année-là. On peut faire l'hypothèse que les enseignants ont adapté les notes aux nouvelles exigences. On peut supposer que les changements d'organisation du CO ne portant pas directement sur les pratiques d'évaluation et d'attribution des notes, ils n'ont que peu d'impact sur l'évolution des moyennes trimestrielles.

1. Introduction

La recherche dont il sera question dans ce rapport a pour thème l'évolution des moyennes trimestrielles et se situe donc dans la problématique de l'évaluation scolaire. Avant d'entrer dans le vif du sujet, nous allons rappeler quelques présupposés liés à l'évaluation au moyen d'une brève revue de la littérature. Tout d'abord, l'évaluation scolaire peut être interne (réalisée par les enseignants à l'intérieur des classes) ou externe (effectuée au moyen d'épreuves élaborées par des groupes d'enseignants ou des organismes externes cantonaux, nationaux ou internationaux). Elle peut également avoir des fonctions variées et complémentaires : certificative ou sommative, formative (réguler l'enseignement) ou bilan servant au monitoring ou au pilotage du système. En d'autres termes, elle peut, selon les cas ou à la fois, juger d'un niveau scolaire, servir de pronostic pour la scolarité future, situer les élèves au sein d'une classe, servir à la gestion de la classe ou encore réguler l'enseignement (Bressoux & Pansu, 2003).

Dans le cadre de cette étude, nous nous situons dans un contexte d'évaluation interne et sommative. Les moyennes trimestrielles sont le fruit de travaux d'évaluations plus ou moins nombreux selon la discipline menés par l'enseignant. Il s'agit le plus souvent de travaux écrits plus ou moins longs et reposant sur des parties de programme de longueur variable.

L'attribution de notes est une obligation incontournable du métier d'enseignant qui doit évaluer les apprentissages de ses élèves et en rendre compte. Toutefois, comme l'ont souligné différents auteurs (Merle, 1996, 1998, 2007 ; Bressoux & Pansu, 2003 ; Chevillard, 1986 ; etc.), la notation ne résulte pas d'une simple application d'un barème de correction ni d'une simple activité de « mesurage » des performances effectives (Chevillard, 1986). Elle résulte d'un processus de fabrication. Ainsi, pour Merle (2007), « *il s'agit d'une activité sociale liée à l'apprentissage : l'enseignant ne note pas seulement un devoir, il note aussi le travail réalisé par un élève singulier dans une classe donnée. Le professeur n'évalue pas seulement des compétences, il doit aussi préserver des relations pédagogiques* » (p. 7-8).

De plus, la notation dépasse la mesure d'une compétence, car elle a un impact sur l'investissement de l'élève dans son travail en le motivant ou le démotivant. Ces effets sont variables selon le niveau des élèves : les jugements des enseignants influencent davantage la réussite des élèves faibles que celle des élèves forts (Maddon, Jussim et Eccles, 1997).

L'attribution de notes aux élèves est donc un processus complexe et qui peut donner lieu à un certain nombre de biais ou d'erreurs. D'après Leclercq, Nicaise et Demeuse (2004), il y aurait trois sources d'erreurs : le système scolaire (contexte ou classe), les enseignants (valeurs et normes) et les élèves (caractéristiques individuelles).

Le système scolaire

Ainsi, le système et le contexte scolaires en général ont une influence sur les processus d'évaluation. En effet, la classe dans laquelle se trouve un élève a un effet sur les notes attribuées. On observe un *effet Posthumus*, c'est-à-dire qu'« *avec les mêmes performances et toutes choses égales par ailleurs, un élève est jugé par son professeur comme un "bon élève" dans une classe alors qu'il peut se voir contraint de doubler son année dans une autre. Tout dépend non pas des performances particulières de l'élève dans l'absolu mais de ses performances par rapport à celles de ses condisciples* » (Leclercq, Nicaise et Demeuse, 2004, p. 275).

Cet effet se manifeste de la manière suivante : l'enseignant a tendance à ajuster le niveau de son enseignement et ses appréciations aux performances des élèves. Il y aurait une « *sorte de prototype communément admis comme quoi il existerait peu d'élèves exceptionnels (très faibles ou très brillants) mais beaucoup d'élèves moyens* » (p. 280).

L'enseignant s'adapte à son public d'élèves et adapte le contenu pédagogique des cours en fonction de leur niveau supposé, le plus souvent au niveau moyen de la classe. Par ailleurs, les attentes des enseignants sont variables : elles sont plus élevées pour les classes supposées meilleures.

D'autres biais extérieurs au travail des élèves peuvent également exister, tels que l'ordre de correction des copies : s'il s'agit des premières ou dernières ou si elles suivent une copie très bonne ou au contraire très faible.

L'établissement scolaire dans lequel sont scolarisés les élèves peut également influencer l'attribution des notes, mais le plus souvent de manière moins marquée que la classe ou le maître (Bressoux, 1994). Mentionnons aussi l'effet de la différenciation scolaire, en l'occurrence du regroupement, qui influence l'évaluation de par le statut scolaire qu'il attribue aux élèves et à la classe.

Les enseignants

Les enseignants ont bien sûr également une influence sur l'attribution des notes, étant relativement libres dans le choix des procédures d'évaluation. Comme le rappellent Leclercq, Nicaise et Demeuse (2004), « *le fait de noter un élève est une action proprement rationnelle qui trouve ses fondements [...] à la fois dans les intérêts et les valeurs propres de l'enseignant. Si le jugement professoral peut être dépendant de l'origine sociale des élèves, de leur âge, de leur sexe ou du type d'établissement fréquenté, il l'est tout autant (Merle 1996) d'un ensemble quotidien d'arrangements et de "bricolage" des notes, intentionnels ou non* » (p. 277).

Les arrangements utilisés par l'enseignant peuvent être internes ou externes. Les arrangements internes sont des stratégies permettant de bien fonctionner dans sa classe en incitant les élèves à travailler, en permettant de garantir le calme et obtenir un certain respect (Merle, 2007).

De manière globale, le jugement scolaire a plusieurs fonctions. Pour Chevallard (1986), il peut avoir une fonction de reproduction des hiérarchies sociales en tant que résultat d'un rapport de force entre personnes ou groupes sociaux avec des positions hiérarchiques différentes dans la société (par exemple, l'enseignant et les élèves). Il peut également avoir une fonction didactique de transaction. Chevallard considère l'attribution de notes comme une transaction s'inscrivant dans un processus beaucoup plus large, de « *négociation didactique* » entre l'enseignant et ses élèves. C'est pourquoi les notes ne peuvent pas avoir une moyenne trop haute (afin de maintenir un certain niveau d'exigence), ni trop basse (pour rester crédible). De même, la dispersion ne devrait être trop élevée (problème de gestion des niveaux de connaissance) ni trop faible (pour des raisons de discrimination).

Les arrangements externes concernent les directions d'école, les collègues, les parents ou les autorités scolaires. L'enseignant doit présenter une image de sa classe la plus acceptable possible aux yeux de ces différents acteurs.

D'après Merle (1996), les deux types d'arrangements sont liés. Ainsi une épreuve mal réussie par les élèves non prise en compte dans la moyenne a une double finalité : interne mais également externe.

Une évaluation ne doit pas être trop sévère, car perçue comme injuste, ni trop favorable, car vécue comme laxiste (Perrenoud, 1998). Un équilibre est nécessaire, c'est pourquoi l'enseignant a tendance à garder les mêmes moyennes et distributions de notes (Grisay, 1984 ; Crahay, 1996). La fabrication de l'échelle de notes repose sur certains postulats : par exemple, les moyennes trimestrielles par élève excluent généralement les extrêmes. En effet, l'échec d'un élève peut être attribué à l'enseignant.

A ces deux types d'arrangements s'ajoute une troisième catégorie, peu prise en compte dans les études car plus difficile à saisir : les représentations et valeurs de l'enseignant par rapport à l'évaluation, qui peuvent également entrer en compte dans l'attribution des notes.

Comme le résume Merle (1996), l'évaluation des élèves n'est donc pas une simple apposition d'une note mais relève de processus et de procédures au croisement des contraintes sociales et des biographies des élèves et des maîtres (p. 306).

En d'autres mots, « le processus d'évaluation est dépendant d'un triple rapport entre le professeur et ses élèves, le professeur et ses contraintes externes, et le professeur et son passé, son intériorité, lorsqu'il s'engage personnellement dans son travail » (Leclercq, Nicaise et Demeuse, 2004, p. 279).

Les élèves

Enfin, un autre élément pouvant également influencer le jugement des enseignants relève des élèves à proprement parler et plus particulièrement de leurs caractéristiques intrinsèques : leur origine sociale, leur genre mais également leur apparence physique. Ces caractéristiques donnent lieu le plus souvent à des stéréotypes : par exemple, les filles sont perçues comme meilleures en français, plus scolaires, les garçons sont considérés comme meilleurs en mathématiques, etc. Les études sur la fiabilité des notes (Merle, 1998) ont d'ailleurs montré qu'un stéréotype positif par rapport à un élève pouvait lui faire gagner plusieurs points dans un travail d'évaluation.

Le niveau scolaire antérieur des élèves joue également un rôle. Les attentes des enseignants vis-à-vis des élèves s'apparentent à l'*effet Rosenthal* (prophéties auto-réalisatrices). De plus, le correcteur a tendance à attribuer à un élève le même type de notes que celles acquises auparavant. En 1978, Noizet et Caverni déclaraient déjà : « *il est probable que les premiers indices recueillis, qu'ils soient positifs ou négatifs, vont guider le recueil des indices (...) l'évaluateur cherchant davantage des indices susceptibles de confirmer ses premières inférences que des indices susceptibles de les remettre en question* » (p. 141).

Borko et Cadwell (1982) ont mis en évidence qu'à performances identiques, plusieurs variables avaient un effet significatif sur le jugement de l'enseignant, parmi lesquelles l'attractivité physique, le comportement en classe, l'origine sociale. Par contre, le genre et la situation familiale ont un effet moins marqué et non univoque car dépendant de la discipline.

Selon Bressoux et Pansu (2003), l'attribution des notes suivrait deux grands principes permettant la régulation du fonctionnement de l'évaluateur : la nécessité de discriminer les élèves les uns par rapport aux autres dans une classe déterminée, et celle de centrer la moyenne de la classe autour d'une valeur socialement acceptable pour ne pas paraître trop indulgent ou au contraire trop sévère et ainsi ne pas discréditer l'évaluateur.

Les effets de ces biais (groupe de référence et effet des caractéristiques individuelles) peuvent se cumuler. Pour Kronig (2012), malgré une notation moins sévère dans les classes faibles, on observe aussi un effet négatif par rapport aux élèves de milieux défavorisés : « *Lorsqu'on cumule les facteurs du genre, de l'origine sociale et de la nationalité, la distorsion représente en moyenne une demi-note* » (p. 59).

Pour d'autres auteurs comme Allal et Lafortune (2008), malgré ces différents biais, l'évaluation repose sur un jugement professionnel, résultat d'un processus qui mène à une prise de décision prenant en compte différents éléments liés à l'expertise professionnelle de l'enseignant (expérience et formation). Pour Mottier Lopez (2009), le jugement professionnel ne s'exerce pas seul, mais s'inscrit dans une multiréférentialité (en interaction avec les différentes personnes impliquées par les décisions : enfant, famille, collègues et institution scolaire).

Allal et Mottier Lopez (2008) énoncent un certain nombre de principes qui définiraient le jugement professionnel en éducation. Nous retiendrons notamment le quatrième où le jugement professionnel est considéré comme impliquant une mise en relation entre différents éléments : les exigences du système scolaire, la culture de l'établissement, les valeurs personnelles et les stratégies pédagogiques de l'enseignant. Pour Dompnier, Pansu et Bressoux (2011), l'évaluation est considérée comme une activité de jugement multidéterminée qui serait sujette à un effet de halo (performances dans une discipline influencent le jugement par rapport à une autre), à un effet de contexte (celui de la classe), mais aussi à un effet des stéréotypes et des normes (non seulement le genre et l'origine sociale mais aussi le fait d'être redoublant).

2. Objectifs de la recherche et méthodologie

La recherche dont nous rendons compte dans ce rapport porte sur l'évolution des moyennes trimestrielles de mathématiques, français et histoire en 1^{re} année du CO à Genève. Il s'agit d'un mandat confié par la direction générale de l'enseignement obligatoire (DGEO) au groupe ÉduMétrie. L'étude a été réalisée en collaboration avec le service de la recherche en éducation (SRED).

Elle repose sur les résultats des élèves de la volée 2012 (deuxième année de fonctionnement du nouveau CO) et aussi, à titre de comparaison, sur ceux de la volée 2007³.

Son objectif était de répondre aux questions suivantes :

- ♦ Peut-on mettre en évidence, dans l'ensemble, une tendance à la progression, à la régression ou à la stabilité des notes tout au long de l'année ?
- ♦ Est-on à même de dégager des profils individuels différents traduisant diverses évolutions ?
- ♦ Quel est l'impact sur ces évolutions de variables telles que le niveau initial de l'élève (sa moyenne annuelle en fin de primaire), ses aptitudes (score aux tests de fin de primaire), son affectation à une filière (regroupement), son genre, sa première langue parlée ou sa catégorie socioprofessionnelle ?
- ♦ Les résultats de ces analyses varient-ils selon la branche et selon la volée considérées, la structure et les modalités d'orientation du CO étant différentes en 2007 et 2012 ?
- ♦ Ces résultats peuvent-ils trouver une explication dans l'organisation du CO, dans les caractéristiques des branches ou encore dans les habitudes d'évaluation des enseignants ?

Pour cette étude, la méthode retenue est le *modèle multiniveau de croissance*, approche statistique particulièrement adéquate pour l'analyse de profils d'évolution dans le temps. Ce modèle est, en réalité, une extension des modèles multiniveaux classiques, qui permet plus particulièrement de modéliser les trajectoires de croissance de processus et de tester l'effet de différentes variables explicatives aux différents niveaux de hiérarchie (Leroy et Bressoux, 2008). En effet, les données utilisées possèdent une structure hiérarchisée où des mesures répétées dans le temps constituent un premier niveau d'analyse. Ces mesures sont, par ailleurs, regroupées dans une unité supérieure correspondant à un élève, qui constitue un second niveau (voir chap. 5, *Présentation du modèle multiniveau de croissance*). Dans cette approche méthodologique, la personne constitue, en quelque sorte, « l'environnement » qui va influencer sur la mesure.

³ On s'intéresse à la 7^e de 2007 et à la 9^e de 2012 qui correspondent, toutes les deux, à la 1^{re} année de scolarité du CO.

3. Organisation du rapport

Après une introduction (section 1) et un rappel des objectifs de la recherche et de la méthodologie mise en œuvre (section 2), la section 4 traite de l'échantillonnage des données au sens large en fournissant des éléments sur les volées 2012 et 2007, leurs différences et le choix des disciplines retenues dans l'analyse. La section 5 est consacrée à une présentation plus détaillée du modèle multiniveau de croissance, qui est apparu comme la méthode statistique la plus adaptée pour le traitement de la problématique de l'évolution des résultats trimestriels. La section 6 du rapport se centre sur les résultats de l'analyse : dans un premier temps, ce sont les résultats de la volée de 9^e de 2012 que l'on aborde (résultats de mathématiques, français et histoire ainsi qu'évolution simultanée des moyennes dans ces trois disciplines), puis sont présentés ceux de la volée de 7^e de 2007. Enfin, une 7^e section conclusive rappelle les résultats les plus saillants de l'étude et tente de fournir des perspectives de recherche pour mieux connaître les conditions de l'évolution des notes dans l'enseignement secondaire I.

4. Échantillonnage des données

La volée 2012

Compte tenu de son objet, nous n'avons conservé dans notre étude que les élèves qui fréquentaient pour la première fois la 1^{re} année du CO (9^e CO), qui avaient suivi toute l'année le même regroupement et dont les notes avaient été données en principe par les mêmes enseignants pour les trois trimestres. Ont donc été écartés de notre étude les élèves qui redoublaient leur 9^e en 2012 (n = 124) ; qui ont changé de regroupement en cours d'année (n = 332)⁴ ; qui sont arrivés au CO ou l'ont quitté en cours d'année (n = 26) ; qui n'avaient pas de note au troisième trimestre sur l'ensemble des disciplines (n = 3). Nous n'avons pas non plus pris en compte les classes d'accueil. En ce qui concerne les élèves qui ont changé d'orientation pour rejoindre un regroupement plus exigeant (et n'ont donc pas été retenus dans nos analyses), 87 sont passés de R1 à R2, et 218 de R2 à R3 ; nous en tiendrons compte dans l'interprétation des résultats⁵. Au total, la volée analysée comprend 3'638 élèves, soit le 88% de l'ensemble des élèves de 9^e CO de l'année scolaire 2012-2013, scolarisés dans des classes à regroupements, avec une répartition qui ne diffère pas de la population totale (**Figure 1**).

Figure 1. Année scolaire 2012-2013 : effectif total de la volée selon le regroupement et effectif de la population analysée

	Effectif	%	Effectif de la population analysée	% de la population analysée
Regroupement 1 (R1)	343	8%	289	8%
Regroupement 2 (R2)	1017	25%	857	24%
Regroupement 3 (R3)	2763	67%	2492	68%
Ensemble	4123	100%	3638	100%

Source : SRED.

La volée 2007

Nous avons appliqué les mêmes critères de sélection et d'exclusion pour la volée 2007. Ont donc été conservés les élèves fréquentant des classes du système à regroupements (A, B et C ; n = 3'288) et les classes hétérogènes (n = 605). Ont été écartés les élèves qui appartenaient à des classes d'accueil ; qui redoublaient leur 7^e CO en 2007 (n = 163) ; qui ont changé de regroupements (n = 84) ou de système⁶ (n = 16) en cours d'année ; qui sont arrivés au CO ou l'ont quitté en cours d'année (n = 25). En ce qui concerne les élèves qui ont changé d'orientation, 65 élèves ont rejoint en cours d'année un regroupement plus exigeant (42 de B en A, 22 de C en B, et 1 de C en A) et 19 ont été dirigés vers un regroupement moins exigeant (17 de A en B, 1 de A en C, et 1 de B en C). Au total, la volée analysée comprend 3'893 élèves, soit 93% de l'ensemble des élèves du 7^e degré de l'année scolaire 2007-2008, scolarisés dans des classes à regroupements ou hétérogènes (**Figure 2**).

⁴ Les passerelles ont été approximées par l'affectation d'un élève à un regroupement au 15 mai 2013, différent de ce qu'il était au 15 octobre 2012.

⁵ 18 élèves ont quitté une classe d'accueil pour une classe ordinaire. Par ailleurs, les transferts intra-annuels vers un regroupement moins exigeant sont marginaux (9 élèves).

⁶ Passage d'une classe hétérogène à une classe à regroupement, ou sortie d'une classe d'accueil.

Figure 2. Année scolaire 2007-2008 : effectif total de la volée selon le regroupement (ou classe hétérogène) et effectif de la population analysée

	Effectif	%	Effectif de la population analysée	% de la population analysée
Regroupement C	120	3%	104	3%
Regroupement B	726	17%	604	16%
Regroupement A	2715	65%	2580	66%
Classes hétérogènes (H)	620	15%	605	16%
Ensemble	4181	100%	3893	100%

Source : SRED.

Différences entre les volées 2007 et 2012

Les réorientations intra-annuelles sont proportionnellement plus nombreuses en 2012 car elles concernent environ 8% des effectifs, particulièrement des jeunes qui rejoignent un niveau scolaire plus exigeant (passerelles promotionnelles) et qui avaient donc un niveau scolaire élevé comparativement à ceux de leur regroupement. En 2007, les réorientations en cours d'année ne concernent que 3% de la volée. A l'instar de la volée 2012, elles sont également plus fréquemment utilisées dans une vision promotionnelle. Pour 2012 surtout, une part des jeunes avec un bon niveau scolaire fréquentant le regroupement 2 (et le 1 dans une moindre mesure) sont donc absents des analyses, ce qui pourrait avoir, à la marge, une incidence dans les différences de notations entre les regroupements.

Le choix des disciplines

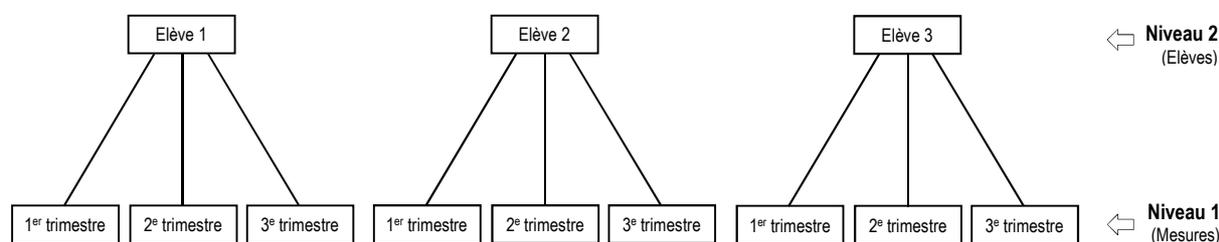
Trois disciplines ont été retenues pour les analyses : les mathématiques, le français et l'histoire. Les mathématiques et le français sont des disciplines fondamentales dans le processus d'enseignement et de sélection du CO ; elles ont notamment un statut particulier dans les normes formelles de promotion et d'orientation⁷. Elles présentent également un aspect cumulatif assez marqué (particulièrement pour les mathématiques) et recouvrent une gamme de compétences larges (particulièrement pour le français). L'histoire est une discipline de culture générale dotée de moins d'heures d'enseignement, dont le caractère cumulatif du plan d'étude est moindre et qui n'a pas le statut des deux autres disciplines pour la promotion et l'orientation des élèves. Considérées ensemble, elles semblent donner une image assez complète du niveau scolaire des élèves et paraissent potentiellement illustratives de la réussite des élèves et de l'évolution de leurs résultats.

⁷ Règlement du cycle d'orientation (RCO) C1 10.26, article 21 alinéa 4.

5. Présentation du modèle multiniveau de croissance

Le modèle multiniveau de croissance, extension des modèles multiniveaux « classiques », apparaît particulièrement bien adapté à la problématique qui nous intéresse dans la mesure où il permet non seulement de tenir compte de la structure hiérarchisée des données (des mesures répétées dans le temps et regroupées dans une unité supérieure correspondant à un élève) (**Figure 3**), mais aussi de modéliser des trajectoires de croissance et de tester l'effet de variables explicatives situées à différents niveaux de la hiérarchie. Ainsi, ce modèle va nous permettre de postuler que la courbe d'évolution des résultats trimestriels (ou courbe de croissance) a une dimension générique, commune à tous les élèves, mais également que chaque élève possède sa propre courbe de croissance qui diffère plus ou moins de la moyenne.

Figure 3. Structure hiérarchisée des données dans la modélisation multiniveau de croissance



L'analyse de données longitudinales à l'aide d'un modèle multiniveau de croissance est une démarche en plusieurs étapes. La première consiste à observer quel est l'impact relatif (en % de la variance) des niveaux définis sur l'évolution des résultats. Dans notre recherche, il s'agit de deux niveaux : d'une part, l'influence du temps sur la variation des moyennes trimestrielles pour un même élève au cours de l'année (profils de progression/régression ; variations intra-individuelles) (niveau 1) ; d'autre part, l'effet des variations imputables aux différences entre élèves (variations interindividuelles) (niveau 2). La seconde étape de l'analyse est l'estimation d'un modèle inconditionnel de croissance qui ne comporte qu'une seule variable explicative, à savoir le temps (variable de niveau 1). Ce dernier est donc considéré ici comme une variable explicative de la moyenne trimestrielle. C'est à ce stade de la procédure que l'on pourra déterminer si le rythme d'évolution des moyennes est bien différent d'un individu à l'autre, autrement dit s'il existe différents types de profils d'évolution des résultats trimestriels chez les élèves. Si tel est le cas, il sera alors possible d'estimer un modèle plus complet, qui permettra notamment de voir dans quelle mesure les différents profils sont liés aux caractéristiques sociodémographiques et scolaires usuelles des élèves (variables de niveau 2). A titre d'illustration de la démarche, l'ensemble de ces étapes sera détaillé ultérieurement pour l'analyse des résultats trimestriels de mathématiques de la volée 2012.

Avant d'aller de l'avant, il convient encore de relever quelques éléments importants sur le modèle de croissance et sa mise en œuvre dans la présente recherche. Tout d'abord, comme nous venons de le mentionner, dans la formulation du modèle multiniveau de croissance, on suppose que le niveau ou le rythme de croissance de la moyenne trimestrielle de mathématiques, français ou histoire est une fonction du temps (le temps est entré dans le modèle comme une variable explicative de la moyenne trimestrielle). Bressoux (2010) relève que, dans ce type de modélisation, le temps peut être défini de différentes manières : en âge, en durée écoulée entre deux mesures, en fonction de la classe fréquentée dans le monde scolaire (par exemple 1^{re} année de scolarité obligatoire, 2^e année de scolarité obligatoire), etc. Dans tous les cas, le choix doit se porter sur la cadence qu'on suppose être la plus judicieuse pour décrire comment évaluer la variable *réponse*. Dans le cadre de notre étude, les données

misés à disposition par la DGEO comportent trois mesures réalisées par les enseignants à la fin de chacun des trimestres de l'année scolaire. La cadence est donc, en quelque sorte, imposée par les données. Par ailleurs, en pratique, il est en général commode de fixer la première valeur du temps à zéro car cela procure une interprétation directe de la constante dans le modèle. En effet, cette dernière correspond alors à la valeur de la variable *réponse* au moment de la première mesure (c'est-à-dire, dans notre cas, la moyenne trimestrielle de mathématiques, français ou histoire à l'issue du 1^{er} trimestre). Pour l'ensemble de ces raisons, la variable *temps* retenue pour nos analyses a été construite de la manière suivante :

$$\text{Temps}_{ti} = \begin{cases} 0 = 1^{\text{er}} \text{ trimestre} \\ 1 = 2^{\text{e}} \text{ trimestre} \\ 2 = 3^{\text{e}} \text{ trimestre} \end{cases}$$

Notons également que les données dont dispose le SRED ne se composent que de trois mesures et qu'il s'agit là du minimum requis pour envisager une analyse de croissance (ceci nous contraint à estimer une fonction linéaire du temps). Le modèle de croissance est, en réalité, très lourd en termes de données puisque non seulement il exige au moins trois mesures par individu, mais en plus chacune d'entre elles est supposée mesurer un même concept. Bressoux (2010) relève que cette contrainte n'est pas toujours facile à atteindre en sciences sociales. Par exemple, mesurer une habileté quelconque nécessite, dans bien des cas, un réétalonnage de la mesure. En effet, on ne va pas confronter aux mêmes écrits des individus de 6, 10, 15 ou 25 ans. Le support écrit devra évoluer en fonction du niveau d'habileté des individus. Mais dès lors, est-ce bien toujours la même habileté que l'on mesure ? En réalité, on se situe rarement dans le cas d'une mesure totalement invariante telle qu'une mesure de longueur, de surface ou de temps (croissance de la taille, du poids, évolution des performances en saut en hauteur, au 1500 m, etc.). Par ailleurs, même lorsqu'il n'est pas nécessaire de modifier les instruments de mesure d'une période sur l'autre, le problème de l'univocité du concept mesuré se pose tout de même : une attitude, une habileté peut se modifier par la simple passation de l'épreuve qui les mesure ; faire passer plusieurs fois une épreuve à des individus peut entraîner des progrès dus à cette simple passation. Dans la présente étude, peut-on considérer que les moyennes trimestrielles dont on dispose mesurent toujours la même chose au cours du temps ? Quel est le sens véritable des évolutions temporelles que l'on observe ? La moyenne trimestrielle de mathématiques, français ou histoire ne constitue bien évidemment pas une mesure invariante des niveaux de connaissances et compétences des élèves. En effet, on peut espérer que ces derniers évoluent au cours de l'année scolaire grâce à l'action de l'enseignant (progression dans les apprentissages). En revanche, on peut faire l'hypothèse que la moyenne trimestrielle représente une mesure relativement invariante de l'adaptation de l'élève aux attentes de l'enseignant ainsi qu'à celles de l'école et à l'avancement dans le programme. Dès lors, la construction d'un modèle de croissance devient envisageable.

6. Présentation des résultats

Les données sur les élèves et l'organisation scolaire dont nous disposons pour étudier l'évolution des moyennes trimestrielles sont extrêmement nombreuses (plusieurs volées et pluralité de disciplines). Le temps et les ressources disponibles, limités, nous ont cependant contraints à restreindre le champ de nos investigations. Dans un premier temps, nous analyserons les résultats trimestriels en mathématiques pour la volée d'élèves ayant fréquenté la 9^e CO en 2012. Nous étudierons plus particulièrement le fonctionnement de la notation dans les trois filières qui composent actuellement l'enseignement secondaire I genevois. Ensuite, nous essayerons de voir dans quelle mesure il existe des convergences ou des divergences pour les moyennes de français et d'histoire. Les résultats sont-ils les mêmes pour ces trois disciplines qui mobilisent divers types de compétences, sont plus ou moins cumulatives et dont le statut scolaire n'est pas identique (les mathématiques et le français sont généralement considérés comme des disciplines « principales », l'histoire plutôt comme une discipline « secondaire ») ? Nous aborderons également la question de l'évolution simultanée des moyennes trimestrielles. Pour un élève donné, les profils d'évolution sont-ils les mêmes en français, mathématiques et histoire ? Enfin, en dernier lieu, nous nous intéresserons aux résultats de la volée d'élèves de 2007 et ses regroupements A, B, C, ainsi qu'aux classes hétérogènes⁸ : l'évolution des résultats trimestriels était-elle différente dans un CO qui n'avait pas la même organisation que celle en vigueur à l'heure actuelle ?

Le cas des moyennes trimestrielles de mathématiques en 9^e CO pour la volée 2012

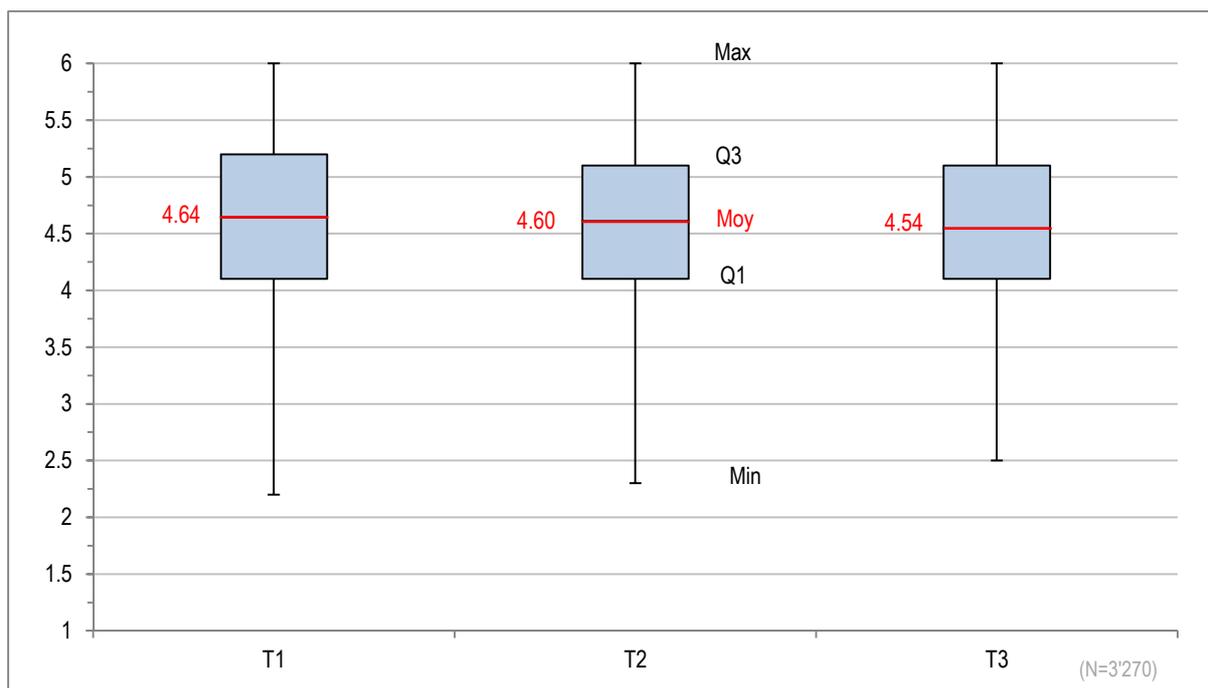
Avant de nous lancer dans l'étude plus complexe des profils d'élèves au moyen de l'analyse multiniveau de croissance, il nous semble opportun de réaliser quelques analyses descriptives à l'aide de méthodes statistiques plus conventionnelles. Rappelons ici, en préambule, qu'au CO les notes attribuées aux différents travaux réalisés par les élèves s'inscrivent sur une échelle de 1 à 6 et qu'une précision supérieure à la fraction $\frac{1}{2}$ n'est pas autorisée (la note de 1 est attribuée à un travail non rendu ou non exécuté). Par ailleurs, le seuil de suffisance est fixé à 4.0 (autrement dit, seules les notes supérieures ou égales à 4.0 sont considérées comme suffisantes). A la fin de chaque trimestre, une moyenne des notes est établie pour chaque discipline sur la base de l'ensemble des divers travaux effectués. Ces moyennes sont arrondies au dixième. Les épreuves communes, qui permettent de situer le niveau de l'élève par rapport aux objectifs d'apprentissage et aux attentes définies dans le plan d'études, entrent dans la composition de la moyenne du troisième trimestre.

Dans un premier temps, nous avons cherché à caractériser l'évolution des moyennes trimestrielles et leur distribution au cours de l'année scolaire de façon générale pour l'ensemble des trois regroupements. Pour ce faire, les résultats trimestriels de mathématiques ont été représentés graphiquement sous la forme de moyennes et de dispersion autour de ces moyennes (**Figure 4**). Il apparaît que les moyennes trimestrielles diminuent très légèrement au cours du temps (on passe de 4.64 au 1^{er} trimestre à 4.54 au 3^e trimestre) et que la dispersion autour de ces moyennes se réduit également. Ainsi, les résultats trimestriels sont légèrement moins bons en fin d'année scolaire, mais sont aussi un peu plus homogènes.

⁸ En 2007, l'enseignement secondaire I genevois est organisé en regroupements assortis d'options. Le regroupement A se caractérise par des exigences scolaires étendues et des effectifs ordinaires : on y trouve les options *langues anciennes* (latin), *sciences* et *arts* (dès la 8^e). Les regroupements B et C (ce dernier pour la 7^e uniquement) correspondent à des exigences scolaires moindres et des effectifs réduits (seules les options *sciences* et *arts* y sont enseignées, dès la 8^e). Notons que le regroupement C n'existe pas dans tous les établissements (il représente seulement 3% des effectifs). Par ailleurs, en parallèle de ce système, trois établissements fonctionnent avec des classes hétérogènes (les élèves n'ont fait l'objet d'aucune sélection sur la base de leurs résultats scolaires antérieurs), des options (langues anciennes, sciences et arts) et des niveaux en allemand et mathématiques.

Figure 4. Tendance centrale et dispersion des moyennes trimestrielles de mathématiques en 9^e CO

(Volée 2012 – Ensemble des regroupements R1, R2 et R3)



Source : SRED.

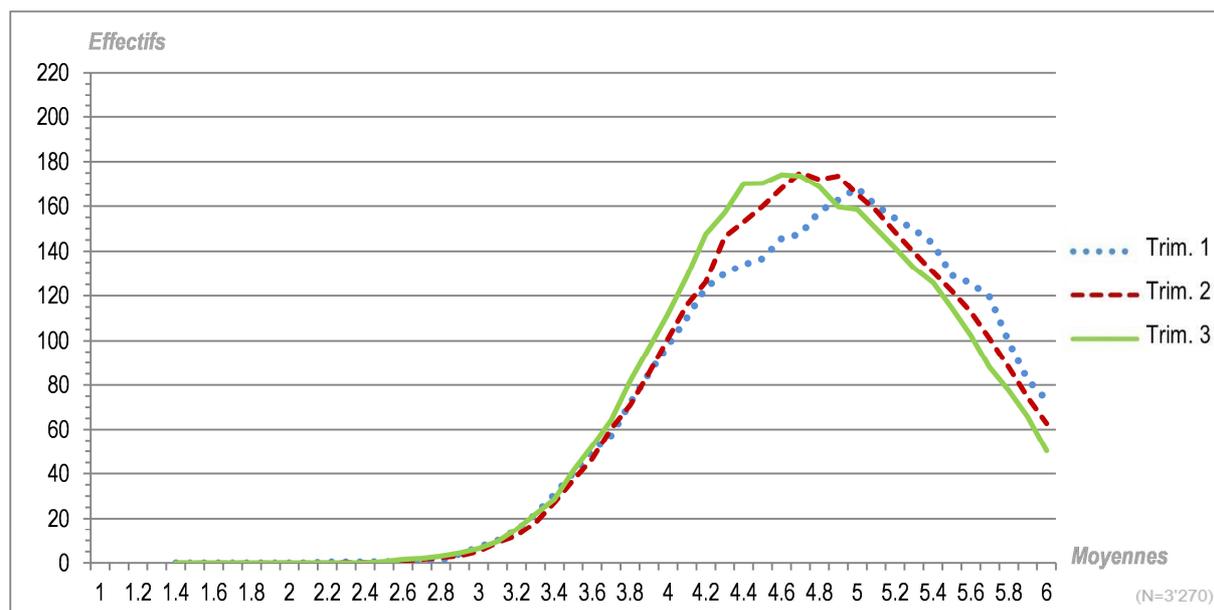
Note de lecture : Les résultats trimestriels de mathématiques sont présentés sous la forme de moyennes et de dispersion autour de ces moyennes à l'aide du graphique connu dans la littérature sous le nom de « box plot » ou « boîte à moustaches ». La longueur de la boîte rend compte de la dispersion des valeurs. De façon générale, la dispersion sera d'autant plus grande que la boîte sera étendue. On rappelle que le 25^e centile (Q1) correspond à la moyenne trimestrielle en dessous de laquelle sont situés les résultats des 25% d'élèves les moins performants. Inversement, le 75^e centile (Q3) correspond à la valeur de la moyenne au-dessus de laquelle sont situés les résultats des 25% d'élèves les plus performants.

Une autre manière de décrire les moyennes trimestrielles de mathématiques en 9^e CO est de représenter leur distribution pour chacun des trimestres (**Figure 5**). On constate tout d'abord que les moyennes sont globalement comprises dans l'intervalle [2.2;6.0]. Par ailleurs, plus de 80% d'entre elles se situent entre 3.8 et 5.6 points. Enfin, on perçoit graphiquement le décalage progressif vers la gauche de la distribution des moyennes entre le premier et le troisième trimestre, ce qui se traduit globalement par une légère diminution de la moyenne des résultats trimestriels au cours de l'année scolaire.

Si l'on considère la variation des résultats entre deux trimestres successifs (**Figure 6**), il apparaît que les évolutions nulles ou relativement faibles sont les plus fréquentes. En effet, en moyenne, l'écart entre deux résultats trimestriels successifs de mathématiques est de -0.05 points (donc très proche de 0). Les différences les plus fréquentes se situent à -0.1 points et 0 points. La distribution des variations n'est, par ailleurs, pas symétrique par rapport à la valeur nulle puisqu'on dénombre plus de valeurs du côté des variations négatives. Enfin, la quasi-totalité (environ 95%) des variations des moyennes trimestrielles de mathématiques sont comprises dans l'intervalle [-0.4;0.3] et 80% des variations sont même comprises dans l'intervalle plus réduit de [-0.2;0.2], ce qui confirme une grande stabilité des notes en cours d'année pour un grand nombre d'élèves.

Figure 5. Distribution des moyennes trimestrielles de mathématiques en 9^e CO selon le trimestre

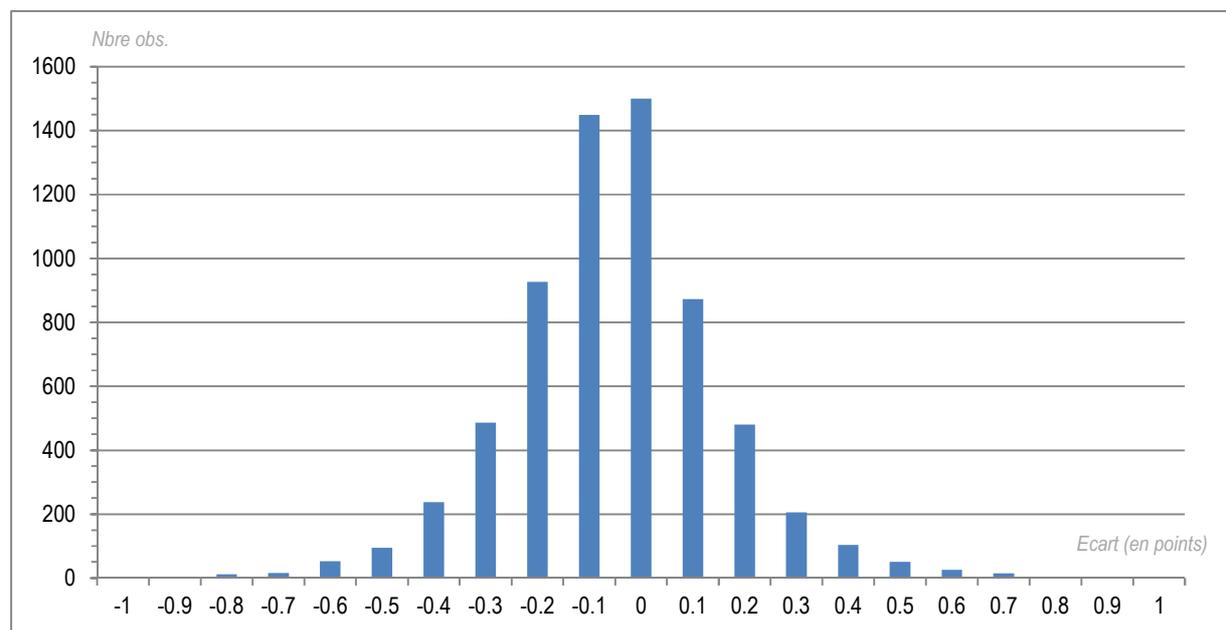
(Volée 2012 – Ensemble des regroupements R1, R2 et R3 – Courbes lissées avec moyenne mobile)



Source : SRED.

Figure 6. Variations trimestrielles des moyennes de mathématiques en 9^e CO

(Volée 2012 – Ensemble des regroupements R1, R2 et R3)

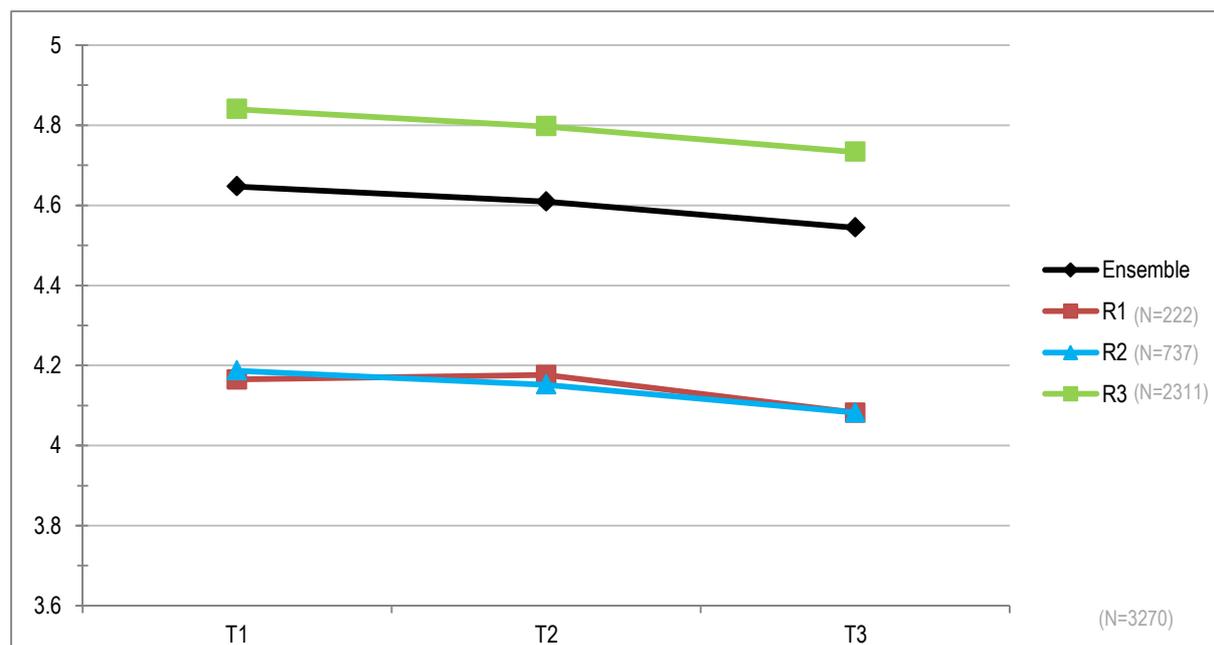


Source : SRED.

Nous avons vu précédemment que les résultats trimestriels de mathématiques diminuent légèrement, en moyenne, au cours de l'année scolaire. Ce constat se retrouve dans chacune des trois filières du CO genevois (**Figure 7**). On observe également que les moyennes dans le regroupement R3, dont les exigences sont les plus élevées, sont systématiquement supérieures à celles des regroupements R1 et R2 qui, par ailleurs, ne se distinguent guère l'un de l'autre. En matière d'évolution des moyennes entre deux trimestres successifs, les variations nulles ou relativement faibles sont les plus fréquentes dans tous les regroupements. On notera également que, dans le regroupement R3, les évolutions égales à 0 ou -0.1 points sont de loin les plus fréquentes, alors que les choses sont un peu plus nuancées en R1 et R2.

Figure 7. Évolution des moyennes trimestrielles de mathématiques en 9^e CO selon le regroupement

(Volée 2012 – Ensemble des regroupements R1, R2 et R3)



Source : SRED.

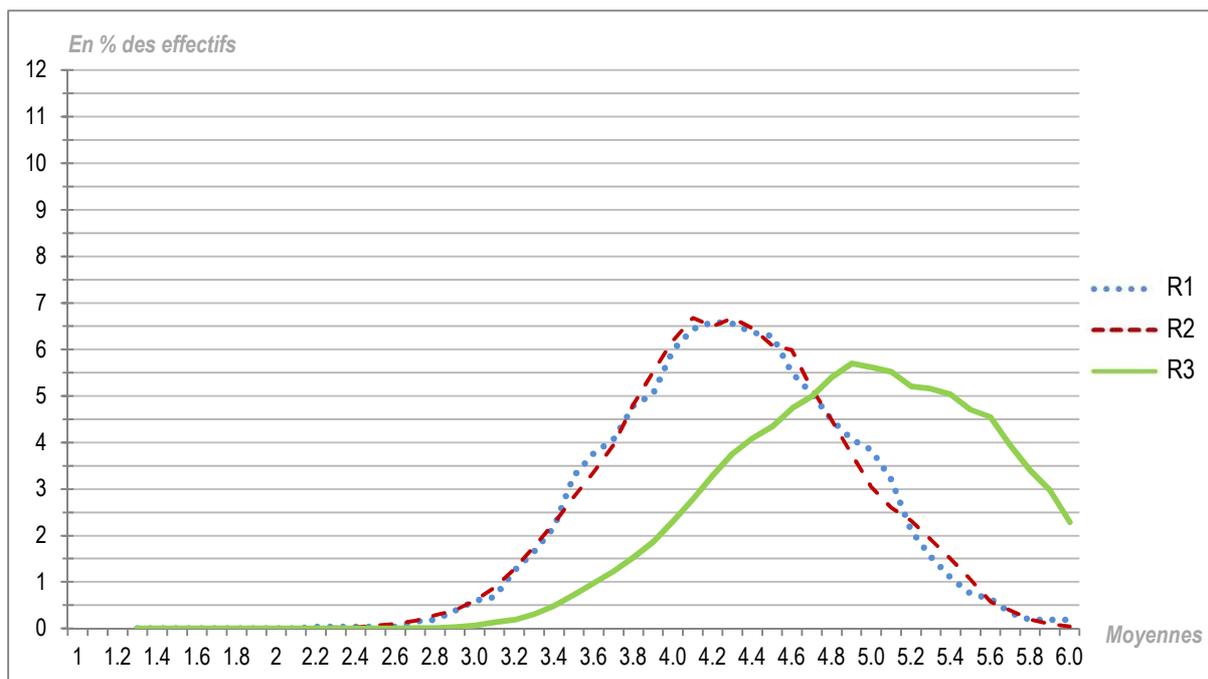
Au-delà de la simple valeur moyenne des résultats trimestriels de mathématiques, l'analyse de la distribution des notes dans les regroupements du CO permet d'illustrer des différences de notation entre les filières (**Figure 8**). Concernant les regroupements R1 et R2, on relève le caractère gaussien, symétrique des distributions et, par-dessus tout, la superposition des courbes qui semble indiquer qu'il n'existe pas de réelle différence dans la façon de noter au sein de ces deux filières. On peut peut-être expliquer cela par le statut scolaire de regroupements se définissant a priori comme recrutant des élèves qui manifestent quelques difficultés dans certains domaines. Inversement, leurs collègues de R3 hésitent plus souvent à considérer comme insuffisants des élèves a priori de bon niveau puisqu'ils ont passé avec succès la première phase de sélection en fin de primaire. Ainsi, la distribution des résultats trimestriels en R3 présente une forme asymétrique, avec une proportion plus importante d'élèves aux notes élevées. Cela se traduit, sans surprise, par une moyenne des résultats trimestriels supérieure à celle observée dans les autres regroupements. On observe donc une différence d'utilisation de l'échelle des notes selon la filière : les enseignants attribuent moins de bonnes notes en R1 et R2 qu'en R3. On peut y voir notamment un effet en retour des dispositions règlementaires relatives aux réorientations (art. 4.3, al. 2). Celles-ci prévoyaient en 2012-2013 la possibilité du passage dans une section aux attentes immédiatement plus élevées en fin de 9^e si la moyenne générale était d'au moins 4.8. En R1 et R2, en attribuant une note égale ou supérieure à ce seuil, les enseignants pourraient craindre d'inciter ainsi l'élève ou/et ses parents à demander un transfert promotionnel. En outre, les objectifs et attentes du plan d'études étant peu précis ou opérationnalisés pour chaque filière, les enseignants de R1 et R2 pourraient avoir tendance à prendre implicitement et partiellement comme référence dans leurs évaluations le niveau de performance attendu ou atteint habituellement en R3⁹.

⁹ En 2012, environ 44% des enseignants de mathématiques et de français qui enseignent en R1 et/ou R2 enseignent également en R3 ; pour les enseignants d'histoire, ce pourcentage est de 60%.

Enfin, en dernier lieu, pour avoir une idée des différents types d'évolution des résultats de mathématiques, on peut étudier les différents types de profils individuels. Nous disposons de trois mesures des résultats de mathématiques au cours de l'année scolaire (une par trimestre). Dès lors, il existe, au total, neuf profils possibles d'évolution des moyennes (résultats constants à chaque trimestre, profil en U, profil en \cap , etc.) (**Figure 11**). Lorsqu'on ne s'intéresse qu'à l'évolution des moyennes entre le premier et le troisième trimestre, on peut réduire l'information à trois profils. On constate qu'une majorité d'élèves (57.2%) présentent des résultats qui ont tendance à diminuer (**Figure 9**). Ceci explique la légère baisse globale des valeurs moyennes mentionnée précédemment.

Figure 8. Distribution des moyennes trimestrielles de mathématiques en 9^e CO selon le regroupement

(Volée 2012 – Ensemble des trimestres – Courbes lissées avec moyenne mobile)



Source : SRED.

Figure 9. Répartition des élèves de 9^e CO selon le profil d'évolution des moyennes trimestrielles de mathématiques

(Volée 2012 – Ensemble des regroupements R1, R2 et R3)

	Effectifs	%
Pas d'évolution des résultats entre le 1 ^{er} et le 3 ^e trimestre	454	13.9%
Progression des résultats entre le 1 ^{er} et le 3 ^e trimestre	942	28.9%
Diminution des résultats entre le 1 ^{er} et le 3 ^e trimestre	1'869	57.2%
Ensemble volée 2012 de 9 ^e du cycle d'orientation	3'265	100%

Source : SRED.

Que nous enseigne l'analyse multiniveau de croissance ?¹⁰

La première étape de l'analyse de croissance, appelée modèle « vide », permet d'évaluer l'importance respective des deux niveaux qui structurent les données (en % de la variance totale) : le niveau des mesures (progressions des notes trimestrielles ; variance intra-individuelle) et celui des élèves (différences interindividuelles). Par ailleurs, ce modèle servira aussi de référence dans la suite des analyses, notamment pour le calcul du pouvoir explicatif du modèle puisque l'évolution de la part de variance résiduelle au fil des spécifications devra toujours lui être rapportée. L'expression détaillée du modèle « vide », qui ne comporte aucune variable explicative, est la suivante :

Au niveau 1 (mesures) :

$$Ma_{ti} = \pi_{0i} + e_{ti} \quad \text{avec} \quad e_{ti} \sim N(0, \sigma_e^2) \quad (1)$$

Au niveau 2 (élèves) :

$$\pi_{0i} = \gamma_{00} + u_{0i} \quad \text{avec} \quad u_{0i} \sim N(0, \sigma_{u0}^2) \quad (2)$$

où :

- t = indice qui désigne les prises de mesure réalisées aux différents temps (varie de 0 pour le premier trimestre à 2 pour le troisième trimestre)
- i = indice qui désigne les élèves
- Ma_{ti} = moyenne de mathématiques obtenue au trimestre t par un élève i
- σ_e^2 = variance intra-individuelle
- σ_{u0}^2 = variance interindividuelle

Lorsque l'on remet les équations (1) et (2) ensemble, on obtient un modèle qualifié de « complet » dont l'expression est la suivante :

$$Ma_{ti} = \gamma_{00} + u_{0i} + e_{ti} \quad \text{avec} \quad e_{ti} \sim N(0, \sigma_e^2) \quad \text{et} \quad u_{0i} \sim N(0, \sigma_{u0}^2) \quad (3)$$

L'estimation de ce modèle, et plus particulièrement des paramètres de variance σ_e^2 et σ_{u0}^2 , est fournie dans la **Figure 10**. Bien que la grande majorité de la variance totale des résultats de mathématiques soit constituée par des différences interindividuelles, le modèle multiniveau de croissance révèle que près de 7% de cette variance correspond tout de même à des variations des moyennes enregistrées d'un trimestre sur l'autre pour un même élève. Un élève est donc bien confronté à des fluctuations de ses résultats scolaires, sa moyenne trimestrielle de mathématiques pouvant être plus ou moins élevée que la précédente et/ou que la suivante (autrement dit, les résultats trimestriels ne sont pas constants dans le temps).

Figure 10. Estimation des paramètres de variance du modèle « vide »

(Volée 2012 – Ensemble des regroupements R1, R2 et R3)

		Estimation des paramètres		Décomposition de la variance
		valeur	signif.	
Niveau 1 (mesures)	Variance intra-individuelle σ_e^2	0.03129		6.51%
Niveau 2 (élèves)	Variance interindividuelle σ_{u0}^2	0.44940	<0.001	93.49%
Variance totale de Ma_{ti}		0.48069		100%

Source : SRED.

¹⁰ Pour une bonne compréhension de ce qui suit, une étude détaillée des formules et des éléments qu'elles comportent n'est pas indispensable.

A l'issue de cette première étape d'analyse, on met en œuvre un second modèle qui intègre le temps comme variable explicative (on parle alors de modèle inconditionnel de croissance). La formulation de ce modèle est la suivante :

Au niveau 1 (mesures) :

$$Ma_{ti} = \pi_{0i} + \pi_{1i} \text{Temps}_{ti} + e_{ti} \quad \text{avec} \quad e_{ti} \sim N(0, \sigma_e^2) \quad (4)$$

Au niveau 2 (élèves) :

$$\pi_{0i} = \gamma_{00} + u_{0i} \quad \text{avec} \quad u_{0i} \sim N(0, \sigma_{u0}^2) \quad (5)$$

$$\pi_{1i} = \gamma_{10} + u_{1i} \quad \text{avec} \quad u_{1i} \sim N(0, \sigma_{u1}^2) \quad (6)$$

Modèle complet (on remplace (5) et (6) par leur expression dans (4))

$$Ma_{ti} = \gamma_{00} + \gamma_{10} \text{Temps}_{ti} + (u_{0i} + u_{1i} \text{Temps}_{ti} + e_{ti}) \quad (7)$$

où :

- π_{0i} = constante de la droite qui décrit l'évolution temporelle pour un individu i , c'est-à-dire son niveau initial (autrement dit la valeur de la variable *réponse* au temps $t=0$)
- π_{1i} = pente de la droite qui décrit l'évolution temporelle pour un individu i , c'est-à-dire son rythme d'évolution
- γ_{00} est une constante moyenne (niveau initial moyen)
- γ_{10} est une pente moyenne (rythme d'évolution moyen).

Envisager la pente de la droite qui décrit l'évolution temporelle de la moyenne trimestrielle de mathématiques pour un individu i (6) comme une variable aléatoire de la forme $\pi_{1i} = \gamma_{10} + u_{1i}$ avec $u_{1i} \sim N(0, \sigma_{u1}^2)$, soit un effet fixe moyen γ_{10} plus un terme aléatoire u_{1i} , permet de postuler que le rythme d'évolution de la moyenne trimestrielle est différent d'un élève à l'autre (autrement dit qu'il existe des « profils » d'évolution différents). Cette hypothèse sera vérifiée dans le cas où le paramètre de variance σ_{u1}^2 est statistiquement significatif (donc différent de 0). Si ce n'est pas le cas, alors l'effet du temps est un effet fixe et on ne peut pas conclure à l'existence de « profils ». Les résultats de cette étape de l'analyse sont présentés dans la **Figure 13** et nous permettent de faire deux nouveaux constats : tout d'abord, le coefficient estimé associé au temps, qui représente le rythme d'évolution moyen pour un élève (autrement dit, qui décrit l'évolution temporelle de la moyenne trimestrielle de mathématiques), est statistiquement significatif et négatif. Cela confirme que les résultats trimestriels de mathématiques ont bien tendance, en moyenne, à diminuer au cours de l'année scolaire. Par ailleurs, on peut aussi montrer que le rythme d'évolution des notes trimestrielles de mathématiques est différent d'un individu à l'autre : il existe donc bien différents types de profils d'évolution des moyennes trimestrielles chez les élèves¹¹.

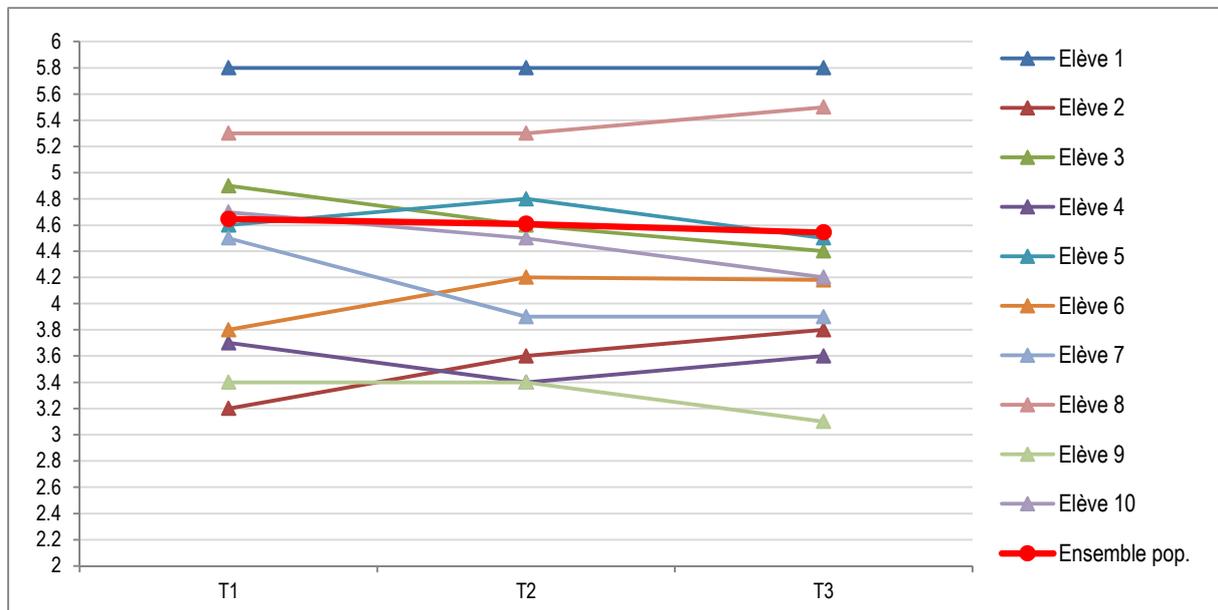
Afin de mieux percevoir le principe de l'analyse multiniveau de croissance, il est possible de proposer une illustration graphique de la démarche de modélisation et des résultats qui précèdent. Dans un premier temps, chaque élève peut être représenté par un nuage de trois points dans un espace où les résultats trimestriels de mathématiques sont une fonction du temps (1 point par trimestre). En reliant chacun de ces points par des droites, on obtient des profils observés d'évolution des moyennes trimestrielles. Ils sont représentés sur la **Figure 11** pour un échantillon d'une dizaine d'élèves. En suivant le même principe, on peut également représenter sur ce graphique l'évolution de la moyenne de l'ensemble de la volée de 9^e de 2012 (en rouge).

¹¹ Le paramètre de variance associé à la pente de la variable *temps* est égal à 0.01589 et il est statistiquement significatif (p-value < 0.001).

On constate que les profils observés peuvent avoir des formes assez différentes, traduisant des évolutions différenciées d'un élève à l'autre : croissance ou décroissance des résultats à chaque trimestre (élèves 2 et 3), résultats constants sur toute l'année scolaire (élève 1), progression de la moyenne entre les deux premiers trimestres puis stabilité des résultats au troisième trimestre (élève 6), etc.

Figure 11. Profils d'évolution des moyennes trimestrielles de mathématiques en 9^e CO - Observations

(Volée 2012 – Ensemble des regroupements R1, R2 et R3)



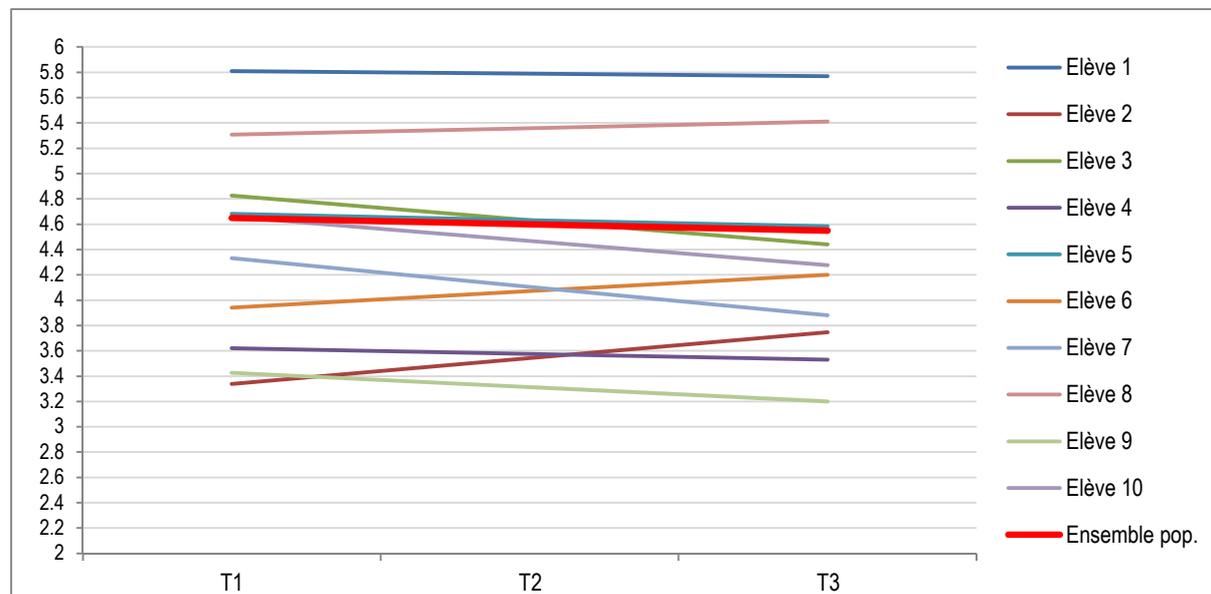
Source : SRED.

Essayer de comprendre l'évolution des résultats trimestriels à l'aide du modèle multiniveau de croissance revient à élaborer une représentation schématique qui s'efforce de reproduire les propriétés fondamentales de l'objet étudié (Bressoux, 2010). Le modèle, bien qu'étant une simplification d'une réalité plus complexe, doit permettre de dépasser le stade de la simple observation pour accéder à une certaine généralité ou universalité. Comme l'indiquait P. Bressoux (2004) au sujet du modèle multiniveau, « les modèles statistiques tels que ceux utilisés dans notre exemple de recherche sont des instruments (cognitifs) fondamentaux au sens où, comme la lunette des astronomes (instrument concret), ils permettent de percevoir la réalité au-delà de ce que nous permettent nos seuls sens » (p. 72). Dans la modélisation de croissance, chacun des profils observés sur la **Figure 11** va être approximé par une droite. Ainsi la **Figure 12** représente, pour la même dizaine d'élèves, l'évolution estimée des moyennes trimestrielles obtenue à l'aide du modèle multiniveau de croissance. On constate que ce dernier permet non seulement de mettre en évidence une dimension générique de la courbe de croissance des résultats trimestriels, commune à tous les élèves (la droite rouge qui indique qu'en « moyenne » les résultats trimestriels diminuent au cours du temps), mais aussi le fait qu'il existe différents types de profils d'évolution. Ceci se traduit graphiquement par l'existence d'une droite de régression pour chaque élève, chacune de ces droites étant différente des autres par sa constante (le niveau initial de l'élève au 1^{er} trimestre) et sa pente (le rythme d'évolution des résultats dans le temps).

Le modèle inconditionnel de croissance nous a permis de constater qu'il existe bien différents types de profils d'évolution des moyennes trimestrielles, mais comment expliquer que le rythme d'évolution des résultats ne soit pas le même pour tous les élèves ? Est-ce lié au fait que ces derniers ne possèdent pas les mêmes caractéristiques individuelles, plus particulièrement sur le plan sociodémographique et scolaire ? La construction d'un modèle de croissance plus complet, que l'on appelle aussi modèle conditionnel de croissance, va permettre de répondre en partie à ces interrogations. En effet, cette modélisation permet de tester l'existence d'un effet des caractéristiques individuelles sur le rythme d'évolution moyen des notes trimestrielles de mathématiques.

Figure 12. Profils d'évolution des moyennes trimestrielles de mathématiques en 9^e CO - Estimations réalisées avec le modèle multiniveau de croissance

(Volée 2012 – Ensemble des regroupements R1, R2 et R3)



Source : SRED.

La formulation du modèle retenu est la suivante :

Au niveau 1 (mesures) :

$$Ma_{ti} = \pi_{0i} + \pi_{1i} \text{Temps}_{ti} + e_{ti} \quad \text{avec } e_{ti} \sim N(0, \sigma_e^2) \quad (8)$$

Au niveau 2 (élèves) :

$$\pi_{0i} = \gamma_{00} + \gamma_{01}R_{1i} + \gamma_{02}R_{2i} + \gamma_{03}\text{Fille}_i + \gamma_{04}\text{Allophone}_i + \gamma_{05}\text{CSPfavo}_i + \gamma_{06}\text{CSPdéfavo}_i + \gamma_{07}\text{Test8P}_i + \gamma_{08}\text{Math8P}_i + u_{0i} \quad \text{avec } u_{0i} \sim N(0, \sigma_{u0}^2) \quad (9)$$

$$\pi_{1i} = \gamma_{10} + \gamma_{11}R_{1i} + \gamma_{12}R_{2i} + \gamma_{13}\text{Math8P}_i + u_{1i} \quad \text{avec } u_{1i} \sim N(0, \sigma_{u1}^2) \quad (10)$$

Modèle complet (on remplace (9) et (10) par leur expression dans (8)) :

$$Ma_{ti} = \gamma_{00} + \gamma_{01}R_{1i} + \gamma_{02}R_{2i} + \gamma_{03}\text{Fille}_i + \gamma_{04}\text{Allophone}_i + \gamma_{05}\text{CSPfavo}_i + \gamma_{06}\text{CSPdéfavo}_i + \gamma_{07}\text{Test8P}_i + \gamma_{08}\text{Math8P}_i + \gamma_{10}\text{Temps}_{ti} + \gamma_{11}R_{1i} \times \text{Temps}_{ti} + \gamma_{12}R_{2i} \times \text{Temps}_{ti} + \gamma_{13}\text{Math8P}_i \times \text{Temps}_{ti} + (u_{0i} + u_{1i} \times \text{Temps}_{ti} + e_{ti}) \quad (11)$$

Avant de commenter les résultats de cette dernière étape de l'analyse qui sont présentés dans la **Figure 13**, il convient de s'arrêter un instant sur la spécification du modèle qui figure ci-dessus. Comme l'indiquent Snijders et Bosker (1999), la spécification d'un modèle est l'une des parties les plus difficiles de l'inférence statistique, car elle repose sur deux leviers principaux qu'il convient de manipuler simultanément : des considérations en lien avec le sujet de la recherche effectuée d'une part et des considérations statistiques d'autre part. Le but de la spécification est de parvenir à un modèle qui décrit correctement les données observées sans complications inutiles. Dans l'analyse multiniveau, dont le modèle de croissance n'est qu'un cas particulier, le nombre de paramètres à estimer est rapidement considérable. Il faut donc être assez vigilant dans la manière de construire le modèle. Comme l'indique Dormont (1999), modéliser exige une certaine parcimonie. Il s'agit bien d'établir une représentation simplifiée d'une réalité nécessairement plus complexe et de construire une approximation acceptable de la relation étudiée. Ces éléments nous ont conduits à adopter l'une des

simplifications proposées par Snijders et Bosker (1999), à savoir que dans les équations de niveau 2, nous n'avons pas systématiquement utilisé le même ensemble de variables pour caractériser l'élève : l'équation (10) ne comporte que le niveau initial de mathématiques (Math8P) et le regroupement de 9^e CO (R1 et R2), alors que l'équation (9) englobe aussi d'autres caractéristiques sociodémographiques et scolaires.

En ce qui concerne les caractéristiques sociodémographiques, il apparaît que seule la catégorie socioprofessionnelle (CSP) possède un effet significatif sur la moyenne de mathématiques et son évolution. Ainsi, « toutes choses égales par ailleurs »¹² (notamment à niveau initial de mathématiques et regroupement d'appartenance en 9^e identiques), on obtient de meilleurs résultats lorsque l'on est issu d'un milieu favorisé et, à l'inverse, de moins bons résultats lorsqu'on appartient aux milieux les plus modestes (par rapport à un élève de CSP moyenne), constat qui rejoint les conclusions de bien d'autres recherches contrôlant cette variable sociale. En revanche, le genre et la première langue parlée n'ont aucun effet sur le résultat trimestriel de mathématiques, une fois contrôlés le niveau initial, l'origine sociale et le regroupement d'appartenance. Comme il s'agit de mathématiques, le facteur verbal joue apparemment un rôle moindre que dans d'autres disciplines ; d'autre part, quand il s'agit de résultats scolaires (notes), la supériorité des performances des garçons, attestée dans d'autres types d'évaluation (certains tests notamment), tend à s'atténuer ou à disparaître.

En matière de caractéristiques scolaires, on relève un effet positif et important du niveau initial de mathématiques (moyenne de 8^e primaire) sur la moyenne trimestrielle en 9^e CO, qui est d'ailleurs le principal facteur explicatif dans les modèles mis en œuvre. « Toutes choses égales par ailleurs » (notamment le regroupement d'appartenance), le résultat trimestriel de mathématiques en 1^{re} année du CO est d'autant plus élevé que le niveau de mathématiques en fin de primaire est élevé. On peut faire un constat relativement similaire avec les tests de préorientation et, plus particulièrement, avec le résultat au test de raisonnement non verbal. « Toutes choses égales par ailleurs » (notamment la note de mathématiques en 8^e et le regroupement d'appartenance en 9^e), la moyenne trimestrielle de mathématiques en 9^e est d'autant plus élevée que le nombre de réponses justes au test de préorientation est important. Comme dans d'autres travaux similaires (Bressoux, 1994 ; Dumay & Dupriez, 2009), on constate que l'enseignement tend à maintenir les différences de performances entre les élèves, pour des raisons qui tiennent probablement à la fois aux aptitudes des élèves et à l'aspect cumulatif des connaissances. On relève également un effet positif d'appartenance au regroupement R1 et, dans une moindre mesure, au regroupement R2 (par rapport au fait d'être scolarisé en R3). Cela signifie qu'un élève obtient une moyenne trimestrielle de mathématiques en 9^e plus élevée lorsqu'il se trouve en R1 et en R2 plutôt qu'en R3. Ce résultat est très probablement le reflet des processus d'évaluation mis en œuvre par les enseignants, plus particulièrement du fait que le niveau d'un élève s'apprécie par rapport à celui de ses pairs, en moyenne, plus faibles en R1 et R2 que dans le regroupement R3.

Le modèle conditionnel de croissance permet enfin de mettre en évidence deux effets d'interaction positifs et significatifs entre le temps et des variables explicatives qui caractérisent l'élève (surligné dans l'équation (11)). On a vu précédemment que les résultats trimestriels de mathématiques ont tendance, en moyenne, à diminuer au cours de l'année scolaire. Le rythme de décroissance des notes est toutefois moins élevé lorsque l'élève est scolarisé dans le regroupement R1 et lorsqu'il possède un niveau initial de mathématiques plus élevé (ces effets d'interaction positifs viennent, en quelque sorte, atténuer l'effet négatif de la variable *temps* qui est le même pour tous les élèves). Tout se passe donc comme si, en R1, les bons élèves en mathématiques étaient plus à même de maintenir leur avantage initial en termes de notation tout au long de l'année, probablement favorisés dans leur évaluation par la comparaison avec leurs camarades moins performants.

A l'issue de la démarche analytique, nous avons cherché à établir le pouvoir explicatif du modèle à chacun des niveaux étudiés. Dans notre modèle, la part de variance expliquée au 1^{er} niveau de l'analyse (intra-individuel) est de 59% et celle au 2^e niveau (interindividuel) est de 42%. Comme c'est souvent le cas dans ce genre d'analyse, le phénomène étudié reste donc partiellement inexpliqué par les variables introduites dans la modélisation. En ce qui concerne le niveau intra-individuel (niveau 1),

¹² Pour être plus précis, il faudrait plutôt dire « toutes choses incluses dans le modèle égales par ailleurs ».

on ne dispose, dans notre étude, d'aucune autre variable explicative que le temps. La prise en compte d'autres variables décrivant un état des individus susceptible de varier d'un trimestre à l'autre aurait sans doute permis de mieux expliquer la variance intra-individuelle (par exemple, les élèves peuvent subir des crises familiales telles qu'un divorce des parents ; il peut aussi s'agir d'états psychologiques, physiques, de transformations liées à l'adolescence, etc.). Au niveau interindividuel, les caractéristiques sociodémographiques et scolaires dont nous disposons expliquent en partie, mais en partie seulement, les profils d'évolution différenciés des élèves (le niveau initial est le principal facteur explicatif). Là encore, l'information que nous avons pu utiliser pour tenter d'expliquer le phénomène reste relativement sommaire dans la mesure où elle est principalement extraite de fichiers de données administratives. D'autres variables, plus directement liées aux processus d'apprentissage ou à d'autres éléments psycho-affectifs (intérêt pour la discipline, motivation, etc.) permettraient vraisemblablement d'affiner les résultats. Par ailleurs, la prise en compte d'éléments relatifs aux processus d'enseignement serait aussi une piste intéressante à explorer. Cela nous conduirait toutefois à réaliser des modèles comportant plus de deux niveaux d'analyse¹³ qui se révèlent rapidement plus complexes à mettre en œuvre et à interpréter.

Figure 13. Modélisation des courbes de croissance en 9^e CO pour les résultats trimestriels de mathématiques

(Volée 2012 – Ensemble des regroupements R1, R2 et R3)

Paramètres	Modèle 1 « vide »			Modèle 2 inconditionnel de croissance			Modèle 3 conditionnel de croissance		
	Coeff.	Standard error	p-value	Coeff.	Standard error	p-value	Coeff.	Standard error	p-value
Effets fixes									
Constante	4.599	(0.011857)	<0.001	4.6503	(0.012359)	<0.001	0.691	(0.118437)	<0.001
Temps				-0.0509	(0.002613)	<0.001	-0.1305	(0.029824)	<0.001
Niveau initial de mathématiques (note 8P)							0.696	(0.005519)	0.008
Résultat test raisonn ^{mt} non verbal 8P (nbre rép. justes)							0.026	(0.002393)	<0.001
Fille							0.019	(0.016837)	0.272
Catégorie socio-professionnelle (réf. CSP moyenne) :									
CSP favorisée							0.113	(0.023061)	<0.001
CSP défavorisée							-0.095	(0.019850)	<0.001
Allophone							-0.033	(0.018454)	0.075
Regroupement de 9 ^e CO (réf. R3) :									
R1							0.781	(0.060004)	<0.001
R2							0.202	(0.034301)	<0.001
Temps x Niv. maths 8P							0.015	(0.005519)	0.008
Temps x R1							0.056	(0.016317)	<0.001
Temps x R2							0.016	(0.009581)	0.09
Effets aléatoires									
<i>Niveau 2 (interindividuel [élèves])</i>									
Variance des constantes σ_{u0}^2	0.4494		<0.001	0.48887		<0.001	0.26193		<0.001
Variance des pentes Temps σ_{u1}^2				0.01589		<0.001	0.01582		<0.001
<i>Niveau 1 (intra-individuel [mesures])</i>									
Variance σ_e^2	0.03129			0.01283			0.01283		
Déviance	6'239.77			4'474.24			2'211.77		

Source : SRED.

¹³ Les enseignants ou les classes pourraient, par exemple, constituer un troisième niveau d'analyse.

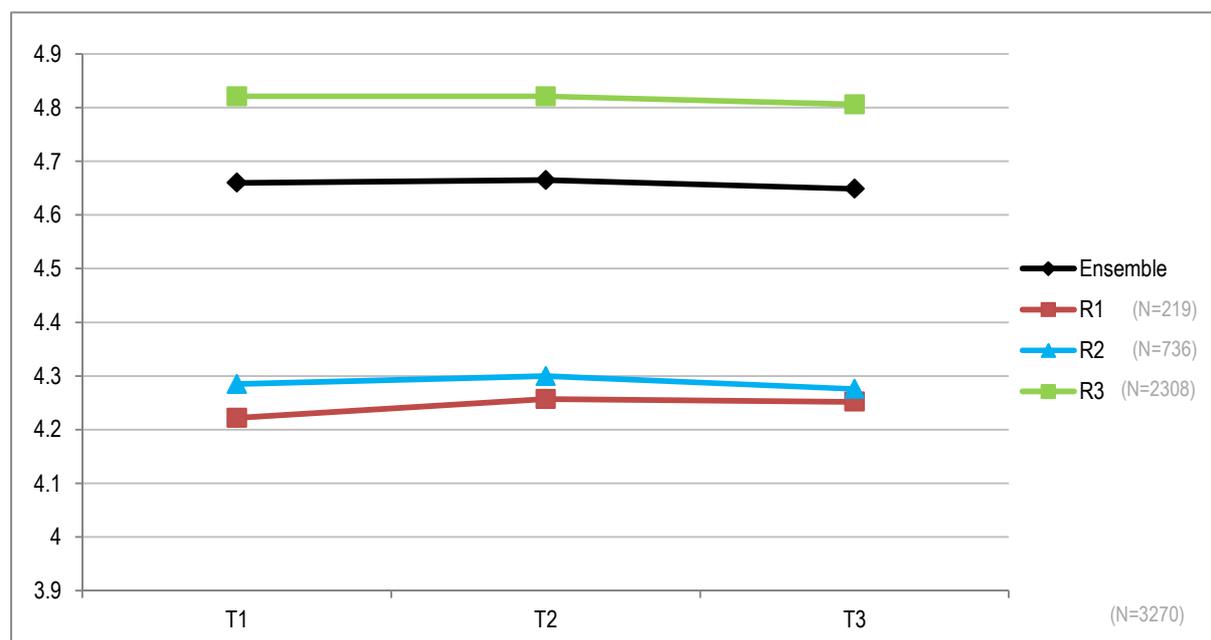
Le cas des moyennes trimestrielles de français et d'histoire en 9^e CO pour la volée 2012

Après avoir étudié en détail les résultats trimestriels de mathématiques et leur évolution, il nous a semblé intéressant de réaliser des analyses similaires pour des disciplines qui mobilisent d'autres types de compétences et/ou qui n'ont pas nécessairement la même importance dans le processus d'orientation (disciplines « principales » et « secondaires »¹⁴). Ainsi, les moyennes trimestrielles de français et d'histoire ont, elles aussi, fait l'objet d'une analyse approfondie reposant à la fois sur des méthodes de statistique descriptive et sur des modèles multiniveaux de croissance. A quelques nuances près, on retrouve des résultats extrêmement proches de ceux mis en évidence pour les mathématiques :

- ♦ en français comme en histoire, les résultats trimestriels ont, en moyenne, tendance à diminuer légèrement au cours de l'année scolaire. Par ailleurs, les variations nulles ou très faibles des notes entre deux trimestres successifs restent les plus fréquentes ;
- ♦ en français et en histoire, comme en mathématiques, les résultats trimestriels moyens sont systématiquement plus élevés dans le regroupement R3 que dans les regroupements R1 et R2 qui, par ailleurs, ne se distinguent pas vraiment l'un de l'autre (**Figure 14**) ;
- ♦ les distributions des moyennes trimestrielles de français et d'histoire sont les mêmes dans les regroupements R1 et R2 (**Figure 15**). En revanche, la distribution des résultats trimestriels en R3 est décalée vers la droite. Elle présente, plus particulièrement, une forme asymétrique avec une proportion plus importante d'élèves aux notes élevées. Les explications de ce décalage sont probablement les mêmes que celles avancées pour les mathématiques (cf. supra).

Figure 14. Évolution des moyennes trimestrielles de français en 9^e CO selon le regroupement

(Volée 2012 – Ensemble des regroupements R1, R2 et R3)

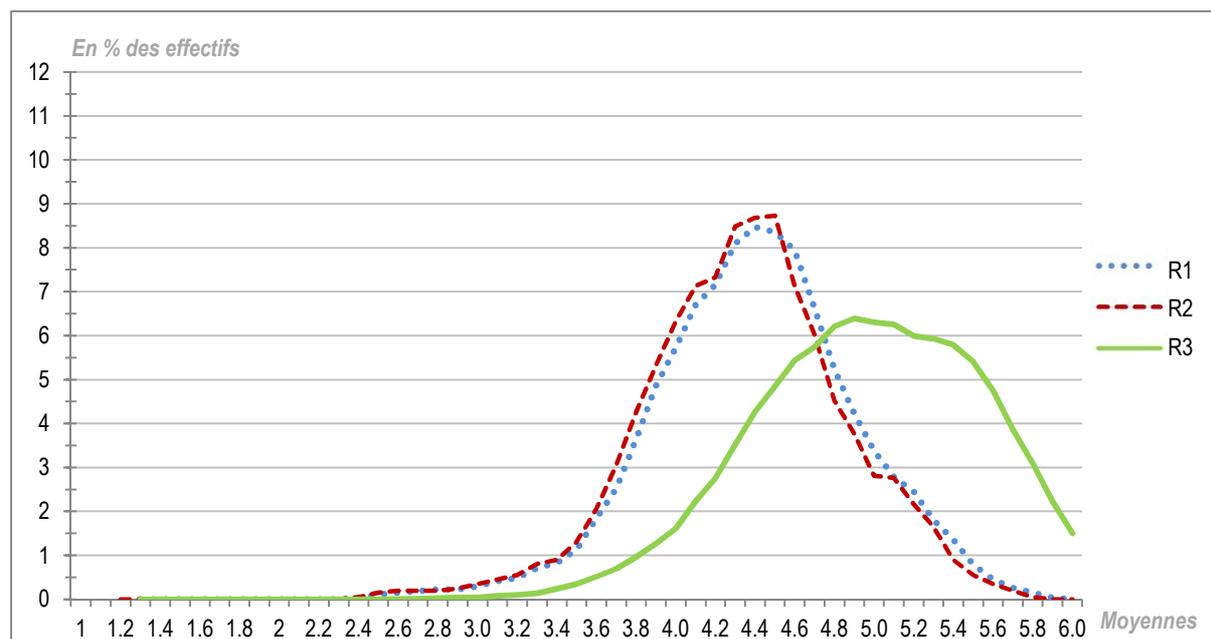


Source : SRED.

¹⁴ Au CO, le français, les mathématiques et l'allemand sont des disciplines principales, reconnues comme fondamentales dans l'apprentissage des élèves et les notes interviennent spécifiquement dans les conditions de promotion et d'orientation.

Figure 15. Distribution des moyennes trimestrielles de français en 9^e CO selon le regroupement

(Volée 2012 – Ensemble des trimestres – Courbes lissées avec moyenne mobile)



Source : SRED.

La mise en œuvre des modèles multiniveaux de croissance nous révèle les éléments suivants (**Figures 16 et 17**).

- ♦ Comme pour les mathématiques, la dimension temporelle n'est pas prépondérante dans la compréhension du phénomène étudié dans la mesure où les parts de variance intra-individuelle (niveau 1) sont faibles (respectivement de 6% et 10.6% pour le français et l'histoire). Ces variances, bien que relativement faibles, sont toutefois statistiquement significatives et il existe donc bien, pour un élève donné, des fluctuations de résultats dans les disciplines concernées au cours de l'année scolaire.
- ♦ Le rythme d'évolution moyen des résultats trimestriels de français et d'histoire pour un élève est statistiquement significatif et négatif. Cela signifie que, comme pour les mathématiques, les résultats trimestriels de français et d'histoire ont tendance, en moyenne, à diminuer au cours de l'année scolaire. Par ailleurs, on peut aussi montrer que le rythme d'évolution des notes est différent d'un individu à l'autre, autrement dit qu'il existe différents types de profils d'évolution des moyennes trimestrielles de français et d'histoire chez les élèves.
- ♦ Comme pour les mathématiques, les différents profils d'évolution des résultats trimestriels de français et d'histoire s'expliquent en partie par la catégorie socioprofessionnelle des élèves. Ainsi, « toutes choses égales par ailleurs » (notamment à niveau initial de français ou d'histoire et regroupement d'appartenance en 9^e identiques), on obtient de meilleurs résultats lorsque l'on est issu d'un milieu favorisé et, à l'inverse, de moins bons résultats lorsqu'on appartient aux milieux les plus modestes (par rapport à un élève de CSP moyenne). On notera aussi que le fait d'être allophone possède un effet négatif sur les résultats trimestriels d'histoire et qu'à l'inverse, le fait d'être une fille possède un effet positif sur les notes de français. Ces deux derniers effets n'avaient pas été mis en évidence dans l'analyse des moyennes de mathématiques. On retrouve ici, comme dans d'autres études analogues, un avantage des filles dans des disciplines qui mobilisent plus particulièrement des aptitudes verbales ainsi qu'un certain handicap pour les élèves dont la première langue parlée n'est pas le français. En ce qui concerne les caractéristiques scolaires, on retrouve l'effet positif du niveau initial et des résultats aux tests de préorientation.

- ♦ On retrouve aussi en français et en histoire le fait que, « toutes choses égales par ailleurs », les élèves de R1 et R2 bénéficient de barèmes plus indulgents.
- ♦ On a vu précédemment que le rythme moyen de décroissance des notes en mathématiques est moins élevé lorsque l'élève est scolarisé dans le regroupement R1 (effet d'interaction Temps x R1 positif et significatif). On retrouve cet effet aussi bien pour le français que pour l'histoire. Par ailleurs, l'effet d'interaction entre le temps et le niveau initial observé pour les mathématiques disparaît au profit d'un nouvel effet d'interaction significatif entre le genre et le temps. Ainsi, la diminution moyenne des notes trimestrielles de français et d'histoire est moins importante pour les filles, quels que soient leurs caractéristiques sociodémographiques et leur regroupement de 9^e (cet effet d'interaction positif vient, en quelque sorte, atténuer l'effet légèrement négatif de la variable *temps* qui est le même pour tous les élèves).
- ♦ Pour le français et l'histoire, le pouvoir explicatif du modèle multiniveau de croissance se révèle un peu plus faible que pour les mathématiques à chacun des niveaux de l'analyse. Dans tous les cas, le phénomène étudié n'est que partiellement expliqué par les analyses mises en œuvre.

Figure 16. Modélisation des courbes de croissance en 9^e CO pour les résultats trimestriels de français

(Volée 2012 – Ensemble des regroupements R1, R2 et R3)

Paramètres	Modèle 1 « vide »			Modèle 2 inconditionnel de croissance			Modèle 3 conditionnel de croissance		
	Coeff.	Standard error	p-value	Coeff.	Standard error	p-value	Coeff.	Standard error	p-value
Effets fixes									
Constante	4.657	(0.010229)	<0.001	4.662701	(0.010876)	<0.001	0.571	(0.100181)	<0.001
Temps				-0.0053	(0.002235)	0.018	-0.0123	(0.003528)	<0.001
Niveau initial de français (note moy. 8P)							0.747	(0.020692)	<0.001
Résultat test raisonn ^{nt} numérique 8P (nbre rép. justes)							0.019	(0.001866)	<0.001
Fille							0.114	(0.016774)	<0.001
Catégorie socio-professionnelle (réf. CSP moyenne) :									
CSP favorisée							0.050	(0.019022)	0.009
CSP défavorisée							-0.058	(0.016483)	<0.001
Allophone							-0.006	(0.015471)	0.702
Regroupement de 9 ^e CO (réf. R3) :									
R1							0.810	(0.047727)	<0.001
R2							0.332	(0.002820)	<0.001
Temps x Fille							0.009	(0.004474)	0.047
Temps x R1							0.025	(0.011246)	0.025
Temps x R2							0.003	(0.005719)	0.614
Effets aléatoires									
Niveau 2 (interindividuel [élèves]) :									
Variance des constantes σ_{u0}^2	0.33509		<0.001	0.37855		<0.001	0.20857		<0.001
Variance des pentes Temps σ_{u1}^2				0.01141		<0.001	0.01137		<0.001
Niveau 1 (intra-individuel [mesures]) :									
Variance σ_e^2	0.0212			0.00977			0.00977		
Déviante	2'734.32			1'497.53			-861.07		

Source : SRED.

Figure 17. Modélisation des courbes de croissance en 9^e CO pour les résultats trimestriels d'histoire

(Volée 2012 – Ensemble des regroupements R1, R2 et R3)

Paramètres	Modèle 1 « vide »			Modèle 2 inconditionnel de croissance			Modèle 3 conditionnel de croissance		
	Coeff.	Standard error	p-value	Coeff.	Standard error	p-value	Coeff.	Standard error	p-value
Effets fixes									
Constante	4.774	(0.01116)	<0.001	4.804	(0.011999)	<0.001	4.259	(0.055281)	<0.001
Temps				-0.0307	(0.003243)	<0.001	-0.0478	(0.004979)	<0.001
Niveau scolaire initial (score Z moy. EC 8P [Fr1, Fr2, Maths])							0.323	(0.026480)	<0.001
Résultat test raisonnement numérique 8P (nbre rép. justes)							0.022	(0.002572)	<0.001
Résultat test synonymes cachés 8P (nbre rép. justes)							0.012	(0.002124)	<0.001
Fille							0.033	(0.020131)	0.099
Catégorie socio-professionnelle (réf. CSP moyenne) :									
CSP favorisée							0.100	(0.023601)	<0.001
CSP défavorisée							-0.054	(0.021354)	0.011
Allophone							-0.065	(0.020129)	0.001
Regroupement de 9 ^e CO (réf. R3) :									
R1							0.348	(0.060492)	<0.001
R2							0.092	(0.035167)	0.009
Temps x Fille							0.033	(0.006473)	<0.001
Temps x R1							0.032	(0.015424)	0.04
Temps x R2							-0.009	(0.008612)	0.277
Effets aléatoires									
Niveau 2 (inter-individuel [élèves])									
Variance des constantes σ_{u0}^2	0.38852		<0.001	0.45252		<0.001	0.30208		<0.001
Variance des pentes Temps σ_{u1}^2				0.02353		<0.001	0.0232		<0.001
Niveau 1 (intra-individuel [mesures])									
Variance σ_e^2	0.04608			0.02164			0.02164		
Déviance	8'340.54			7'121.61			5'689.33		

Source : SRED.

Évolution simultanée des moyennes trimestrielles de français, mathématiques et histoire en 9^e CO pour la volée 2012

En parallèle à l'étude de l'évolution des notes trimestrielles des élèves dans trois disciplines (français, mathématiques et histoire), il est intéressant de connaître leurs évolutions conjointes, afin de mettre en évidence le degré de covariation des notes scolaires pour ces trois disciplines et savoir si les notes de français, de mathématiques et d'histoire évoluent en moyenne pour les élèves de manière conjointe, divergente ou indépendamment les unes par rapport aux autres.

Le coefficient de corrélation entre les notes du premier et du troisième trimestre, pour chaque discipline, est particulièrement élevé (0.91 en français, 0.91 en mathématique et 0.89 en histoire), indiquant la relative stabilité individuelle des résultats (notes) tout au long de l'année, voire même la très forte probabilité de prédiction de la note de fin d'année à partir de celle du premier trimestre.

Le coefficient de corrélation entre les notes de français, de mathématiques et d'histoire est également assez élevé, comme le montre le tableau ci-après (**Figure 18**).

Figure 18. Corrélations entre les notes de français, de mathématiques et d'histoire au 1^{er} et au 3^e trimestre (volée 2012)

	1 ^{er} trimestre	3 ^e trimestre
Français et mathématiques	0.57	0.68
Français et histoire	0.54	0.71
Mathématiques et histoire	0.57	0.64

Corrélations significatives au niveau 0.01.

Source : SRED.

Non seulement les notes ont une forte tendance à covarier, mais cette covariation est plus forte en fin d'année qu'au début. En d'autres termes, les notes des trois disciplines ont tendance à converger davantage en fin d'année. Par hypothèse, cette convergence peut être due à un affinement de l'évaluation-notation au cours de l'année. En début d'année, les élèves proviennent de plusieurs classes et établissements primaires, leur préparation scolaire est variée et l'adaptation aux exigences du CO se fait progressivement. Aussi les différences interindividuelles peuvent-elles être plus importantes. Au cours de l'année, l'affinement de l'évaluation-notation pourrait faire diminuer les erreurs de mesure et, en conséquence, augmenter les corrélations entre branches. Une autre hypothèse serait qu'au cours de l'année, les élèves développeraient peu à peu une certaine attitude générale à l'égard des apprentissages scolaires (intérêt ou désintérêt croissants), affectant donc les trois disciplines.

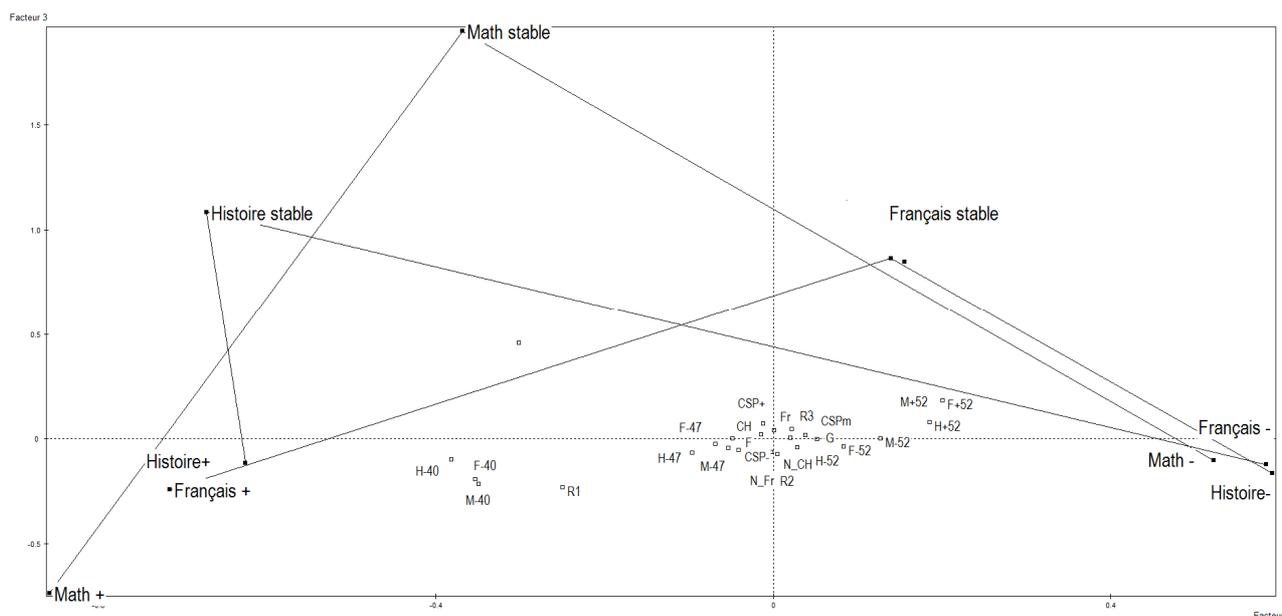
La note du début d'année scolaire est très indicative de celle de fin d'année scolaire, mais la note dans une discipline l'est également (bien qu'à un degré moindre) pour les autres disciplines. Il faut probablement voir derrière ces covariations un socle de compétences scolaires assez largement similaire pour réussir l'ensemble du travail attendu par l'école (ou une absence de ces mêmes compétences qui prêterite l'ensemble du travail). Ce socle de compétences scolaires correspond notamment au fait que dans toutes ces disciplines interviennent des compétences verbales (même en mathématiques) et de raisonnement, ainsi qu'une certaine capacité d'application dans toutes les tâches scolaires. De plus interviennent également de manière transversale le soutien et l'encadrement dont bénéficie (ou pas) l'élève dans le cadre de sa famille. Cette sorte d'aptitude générale à répondre aux exigences scolaires (ou d'inaptitude générale, si les notes sont faibles) semble renforcer son influence durant la première année du CO.

Les corrélations calculées sur les notes de la volée 2007 montrent des covariations très similaires. Les mêmes tendances à la covariation des notes entre trimestres et disciplines s'observent donc pour les deux volées.

Une analyse factorielle des correspondances permet de confirmer ces corrélations en montrant que les profils d'évolution des notes tendent à varier conjointement pour les trois disciplines¹⁵ (**Figure 19**). Le premier axe factoriel (19% de la variance) oppose les élèves qui progressent dans les trois matières à ceux qui régressent dans les trois matières. Ceux qui progressent sont plutôt des élèves qui avaient commencé avec des notes faibles (moins de 4 pour les trois disciplines, voire moins de 4.7 en histoire) en regroupement 1. Ceux qui régressent sont davantage des élèves ayant des très bonnes notes initiales (plus de 5.2 pour les trois disciplines, voire entre 4.7 et 5.2 en français ou en mathématiques) et tendanciellement davantage en regroupement 3. Les variables sociodémographiques ont un impact plutôt faible ; néanmoins, on remarque du côté des progressions plutôt des filles et tendanciellement les élèves de CSP inférieure, et du côté des régressions plutôt des garçons et tendanciellement un peu plus de CSP moyenne. Le troisième axe (17% de la variance également) est essentiellement structuré par les élèves dont les notes sont stables sur l'année ($t_3 = t_1$). Il s'agit plutôt des élèves du regroupement 3, francophones, des classes moyennes et supérieures, ayant des bonnes notes de départ (plus de 5.2 et particulièrement pas en-dessous de la moyenne).

¹⁵ Les profils d'évolution des notes durant l'année ont été définis en trois modalités : pas d'évolution entre le 1^{er} et le 3^e trimestre, progression de la note entre le 1^{er} et le 3^e trimestre, régression de la note entre le 1^{er} et le 3^e trimestre. Quelques autres variables ont été incluses dans l'analyse : la note du 1^{er} trimestre pour des trois disciplines, le regroupement, le sexe, la CSP, la nationalité et la première langue parlée.

Figure 19. Projection factorielle des profils de notes en français, mathématiques et histoire



Source : SRED.

Clé de lecture : Projection des axes 1 et 3 (19% et 17% de la variance). Les profils des notes sont indiqués avec un signe « - » en cas de régression entre le 1^{er} et le 3^e trimestre, avec un signe « + » en cas de progression ; « stable » signifie une note du 3^e trimestre égale à celle du 1^{er}.

Pour les variables supplémentaires ajoutées au modèle, les catégories et leurs abréviations sont les suivantes.

Niveau initial en français : Moins de 4 [F-40], moins de 4.7 [F-47], moins de 5.2 [F-52], 5.2 ou plus [F+52].

Niveau initial en mathématique : Moins de 4 [M-40], moins de 4.7 [M-47], moins de 5.2 [M-52], 5.2 ou plus [M+52].

Niveau initial en histoire : Moins de 4 [H-40], moins de 4.7 [H-47], moins de 5.2 [H-52], 5.2 ou plus [H+52].

Regroupement : R1, R2 et R3

Genre : Garçons [G], filles [F].

Nationalité : Suisses [CH], étrangers [N_CH].

Première langue parlée : Français [FR], autres langues [N_FR].

Catégorie socioprofessionnelle : Supérieure [CSP+], moyenne [CSPm], inférieure [CSP-].

La description des axes est fondée sur les contributions des variables actives et les valeurs-test des variables illustratives (hors de l'intervalle -1.96;1.96).

Une autre organisation du CO : le cas des moyennes trimestrielles en 7^e CO pour la volée 2007

Un autre volet de la présente recherche consistait à comparer l'évolution des moyennes trimestrielles entre les volées 2007 et 2012 afin de voir dans quelle mesure les résultats sont liés, ou non, à l'organisation du CO. En effet, ces deux volées d'élèves ont fréquenté un CO dont l'organisation n'était pas la même (réforme en 2011). De manière résumée, les principales modifications ont consisté à abroger la structure en deux regroupements (A et B) pour les deux dernières années du CO (en 2007, en 1^{re} année ou 7^e CO, il y avait trois regroupements : A, B et C), ainsi que la structure hétérogène avec niveaux et options en vigueur dans trois établissements du canton, pour instaurer dès la rentrée 2011 dans tous les établissements une structure à trois regroupements en 1^{re} année ou 9^e CO (R1, R2 et R3) et trois sections en 10^e et 11^e CO (communication et technologie [CT], langues vivantes et communication [LC], littéraire-scientifique [LS]), ces regroupements et sections correspondant à des

niveaux différents d'exigences scolaires¹⁶. Les conditions d'accès aux trois regroupements en 9^e, basées sur les moyennes annuelles en français I, français II et mathématiques obtenues lors de la dernière année de l'enseignement primaire (8P) ont été réévaluées et sont devenues plus exigeantes qu'auparavant. L'entrée dans les différents regroupements du CO est donc devenue plus sélective, mais parallèlement l'organisation des parcours au sein du CO a été conçue avec un système de passerelles devant permettre des réorientations dites « promotionnelles » vers un regroupement ou une section plus exigeants, que ce soit en cours d'année scolaire ou entre deux années scolaires.

Par ailleurs, le seuil de suffisance au CO pour la volée 2012 est fixé à 4.0 sur une échelle de 6.0 points (autrement dit, seules les notes supérieures ou égales à 4.0 sont considérées comme suffisantes) alors que, pour la volée 2007, il était de 3.5 points. Enfin, la réforme organisationnelle de 2011 s'est aussi accompagnée de l'introduction progressive du Plan d'études romand (PER). Ce plan d'études, mis en œuvre à l'échelle de la région linguistique, décrit les connaissances et les compétences que les élèves doivent progressivement acquérir au cours de leur scolarité obligatoire. Il constitue une référence permettant aux professionnels de l'enseignement de situer leur travail, la place et le rôle de leur discipline dans le cadre du projet global de formation de l'élève et d'organiser leur enseignement.

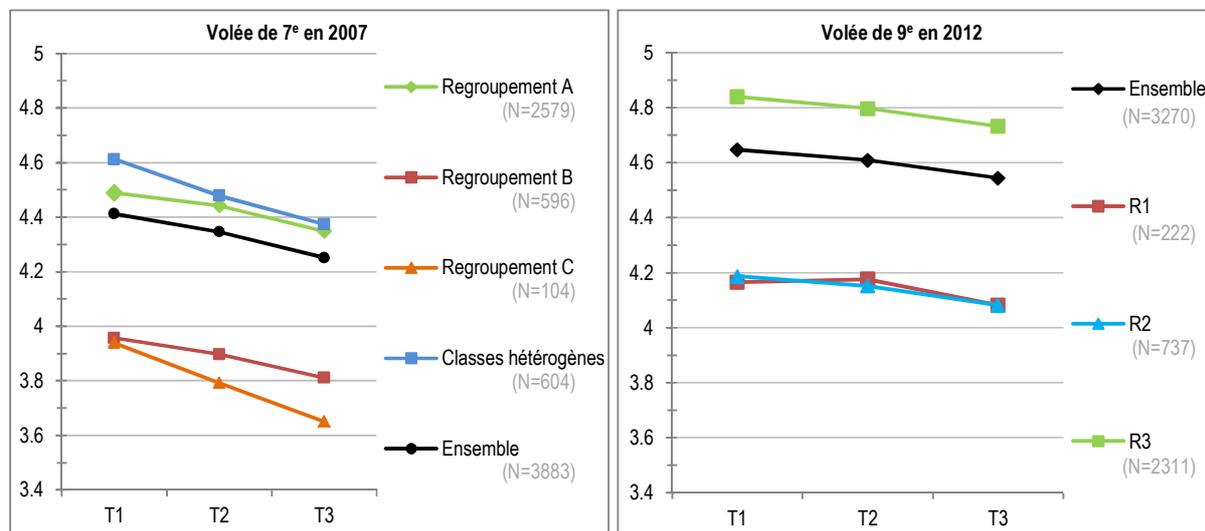
En matière d'évolution des résultats trimestriels, les principaux constats que l'on peut faire sont les suivants :

- ♦ en mathématiques, les moyennes trimestrielles obtenues en 2012 sont systématiquement supérieures à celles obtenues en 2007 (environ 2 à 3 dixièmes en faveur de 2012) (**Figure 20**). On retrouve, par ailleurs, le même résultat pour le français et l'histoire. On peut donc considérer qu'il existe une tendance générale à une utilisation différente de l'échelle des notes dans ces deux organisations du système scolaire. Ce phénomène pourrait être interprété comme la trace d'une modification de la notation des enseignants en 2012 en référence aux nouvelles normes qui balisent l'échelle des notes : les autorités scolaires ont décidé de faire passer le seuil de suffisance de 3.5 à 4.0 entre les deux années considérées ;
- ♦ la comparaison des différentes disciplines entre elles révèle que, en 2007 comme en 2012, les moyennes trimestrielles de mathématiques et de français sont relativement proches et sont inférieures aux moyennes d'histoire (de 1 à 2 dixièmes). Ceci s'explique sans doute par le statut scolaire différencié de ces disciplines qui n'a pas été modifié entre les deux années considérées. En effet, le français et les mathématiques sont, dans les deux cas, des disciplines principales et cumulatives (difficulté progressive) dont les notes interviennent spécifiquement dans les conditions de promotion. L'histoire, en revanche, est une discipline secondaire non cumulative. Le rôle différent joué par les notes de ces disciplines dans l'orientation vers les filières de 2^e année du CO peut laisser supposer que le degré d'exigence des enseignants est plus important dans les disciplines principales. Ceci expliquerait des résultats trimestriels un peu plus faibles en mathématiques et en français ;
- ♦ en comparant les moyennes des regroupements ou filières, on constate pour les trois disciplines et les deux volées un décalage d'environ 6 dixièmes de points entre les moyennes obtenues dans les regroupements ou filières les plus exigeants (A pour la 7^e de 2007 et R3 pour la 9^e de 2012) et les autres (B-C pour la 7^e de 2007 et R2-R1 pour la 9^e de 2012). Ce résultat s'explique non seulement par le statut de ces filières, mais aussi par les références prises par les enseignants pour leurs évaluations : comme en 2012, ils pourraient avoir eu tendance à prendre implicitement et partiellement comme référence dans leurs évaluations le niveau de performance attendu ou atteint habituellement dans la filière la plus exigeante, ce qui tendrait à faire diminuer les notes octroyées dans les autres filières ;
- ♦ pour la volée 2007, on relève que les moyennes trimestrielles de mathématiques (**Figure 20**), de français et d'histoire obtenues dans les classes hétérogènes sont égales ou légèrement supérieures à celles obtenues dans le regroupement A, qui est pourtant la filière aux exigences les plus élevées. Ce constat d'une relative proximité des résultats trimestriels dans ces deux filières pourrait

¹⁶ R1 et CT : exigences élémentaires ; R2 et LC : exigences moyennes ; R3 et LS : exigences étendues. La section LS est par ailleurs composée de trois options : latin, langues vivantes et sciences.

s'expliquer par le fait que dans les deux cas, les enseignants ont en réalité affaire à des classes relativement hétérogènes (le regroupement A accueille trois quarts des élèves scolarisés en A-B-C, soit deux tiers des effectifs de la volée de 7^e). Dès lors, l'utilisation de l'échelle des notes et l'ajustement des barèmes seraient relativement similaires dans les deux contextes.

Figure 20. Évolution des moyennes trimestrielles de mathématiques en 1^{re} année du CO, selon la filière, pour les volées 2007 et 2012



Source : SRED.

Pour dépasser les constats fournis par la simple statistique descriptive, des analyses multiniveau de croissance ont également été mises en œuvre sur la volée 2007. Ces dernières ont été réalisées d'une façon très similaire à celle retenue pour la volée 2012, les données disponibles étant extrêmement comparables (niveau initial, caractéristiques sociodémographiques usuelles et regroupement du CO). La seule différence notable réside dans le contrôle du niveau initial des élèves. En effet :

- ♦ pour la volée 2012, on a utilisé la moyenne des notes de dernière année de primaire (8P) octroyées par l'enseignant pour les mathématiques et le français. Pour l'histoire, de telles informations n'étaient pas disponibles et un niveau scolaire initial plus général a été élaboré. Il correspond à un score Z calculé sur la moyenne des scores aux épreuves cantonales de 8P en français I, français II et mathématiques ;
- ♦ pour la volée 2007, les notes de dernière année du primaire (6P) octroyées par l'enseignant n'étaient pas disponibles. Dès lors, ce sont les résultats aux épreuves cantonales (EC) de fin de primaire qui ont été utilisés pour contrôler le niveau initial des élèves (il s'agit des scores obtenus à l'EC de 6P). De façon plus précise, les choix retenus sont les suivants :
 - français : score à l'EC de français de 6P (français I + français II),
 - mathématiques : score à l'EC de mathématiques de 6P,
 - histoire : niveau scolaire initial plus général correspondant au score Z^{17} calculé sur la moyenne des scores aux EC de 6P en français I, français II et mathématiques (même démarche que pour la volée 2012).

¹⁷ Score centré et réduit.

La mise en œuvre du modèle vide et du modèle inconditionnel de croissance pour la volée de 7^e de 2007 donne des résultats très comparables à ceux observés sur la volée de 9^e de 2012. Il semble qu'il existe une certaine robustesse des résultats indépendamment de l'organisation en vigueur au CO. En effet, on retrouve plus particulièrement les éléments suivants :

- ♦ la grande majorité de la variance totale des résultats trimestriels de 2007 est constituée par des différences interindividuelles, mais une faible partie correspond tout de même à des variations des moyennes enregistrées d'un trimestre sur l'autre pour un même élève (environ 6% pour les mathématiques et le français et 11% pour l'histoire) (**Figure 21**). Comme pour l'année 2012, les résultats trimestriels d'un élève de 2007 ne sont donc pas constants dans le temps (autrement dit, un élève est bien, en moyenne, confronté à des fluctuations de ses résultats scolaires) ;
- ♦ en 2007, comme en 2012, les résultats trimestriels de mathématiques, de français et d'histoire ont tendance, en moyenne, à diminuer au cours de l'année scolaire (**Figures 22, 23 et 24**) ;
- ♦ comme pour la volée 2012, le modèle permet de mettre en évidence l'existence de profils d'évolution différents en français, mathématiques et histoire en 2007 (autrement dit, les rythmes d'évolution des moyennes trimestrielles sont différents d'un élève à l'autre).

Figure 21. Comparaison des résultats des modèles vides pour les volées 2007 et 2012

	Volée de 7 ^e CO (2007)						Volée de 9 ^e CO (2012)		
	Variance			Part de variance (en %)			Part de variance (en %)		
	Mesures (niv. 1)	Elèves (niv. 2)	Total	Mesures (niv. 1)	Elèves (niv. 2)	Total	Mesures (niv. 1)	Elèves (niv. 2)	Total
Français	0.02918	0.472035	0.50122	5.8%	94.2%	100%	6.0%	94.0%	100%
Mathématiques	0.04517	0.69242	0.73759	6.1%	93.9%	100%	6.5%	93.5%	100%
Histoire	0.05999	0.52508	0.58507	10.3%	89.7%	100%	10.6%	89.4%	100%

Source : SRED.

La construction de modèles plus complets, qui intègrent les caractéristiques sociodémographiques et scolaires usuelles des élèves, permet de fournir une explication à ces profils (**Figures 22, 23 et 24**). Là encore, à de rares exceptions près, les résultats sont très proches de ceux observés sur la volée 2012. On relève en particulier :

- ♦ un effet positif de la CSP favorisée (par rapport à un élève de CSP moyenne) sur la moyenne trimestrielle de français, de mathématiques et d'histoire et, à l'inverse, un effet négatif de la CSP défavorisée (à l'exception du français). D'une manière générale, en 2007 comme en 2012, « toutes choses égales par ailleurs » (notamment le niveau scolaire initial et le regroupement du CO), on obtient de meilleures notes lorsque l'on est issu d'un milieu favorisé et, à l'inverse, de moins bons résultats quand on appartient à un milieu plus modeste. Quelle que soit l'organisation du CO en vigueur, le milieu social continue à influencer sur la réussite des élèves au secondaire I alors même qu'il est déjà intervenu dans leur sélection à l'entrée dans ce degré d'enseignement ;
- ♦ un effet positif du niveau scolaire initial (6P) sur la moyenne trimestrielle de français, de mathématiques et d'histoire en 7^e CO : « toutes choses égales par ailleurs » (notamment le regroupement d'appartenance), la moyenne trimestrielle est d'autant plus élevée que le niveau scolaire initial est élevé. Par ailleurs, comme en 2012, on note que le niveau initial reste le principal facteur explicatif dans les modèles mis en œuvre. Le constat d'un enseignement qui tend à maintenir les différences de performances entre élèves (importance des prérequis antérieurs dans les apprentissages actuels, constance des aptitudes des élèves, méthodes d'enseignement qui peinent à compenser les « handicaps » de départ) reste avéré dans les deux volées étudiées ;
- ♦ un effet positif du genre (être une fille) sur les résultats trimestriels de français et un effet négatif de la première langue parlée (allophone) sur les résultats trimestriels d'histoire (mêmes résultats en 2007 et 2012) ;

- ♦ un effet positif d'appartenance au regroupement B, au regroupement C et aux classes hétérogènes (par rapport au fait d'être scolarisé en regroupement A, filière aux exigences les plus élevées) : « toutes choses égales par ailleurs » (notamment le niveau initial), la moyenne trimestrielle de français, de mathématiques et d'histoire en 7^e CO est plus élevée lorsque l'on se trouve en regroupement B, en regroupement C et dans les classes hétérogènes plutôt qu'en regroupement A. En matière d'évolution des résultats trimestriels (et seulement d'évolution), il existe donc une situation comparativement plus favorable pour les élèves scolarisés dans les regroupements aux exigences les plus faibles (par rapport à ceux qui auraient des caractéristiques sociodémographiques et scolaires comparables mais qui seraient scolarisés dans une filière aux exigences accrues). L'évaluation d'un élève par rapport à ses pairs qui, en moyenne, sont plus faibles dans le regroupement B, le regroupement C et les classes hétérogènes peut, en quelque sorte, constituer un avantage en termes d'évolution des résultats trimestriels.

Figure 22. Modélisation des courbes de croissance en 7^e CO pour les résultats trimestriels de mathématiques

(Volée 2007 – Ensemble des regroupements A, B, C et classes hétérogènes)

Paramètres	Modèle 1 « vide »			Modèle 2 inconditionnel de croissance			Modèle 3 conditionnel de croissance		
	Coeff.	Standard error	p-value	Coeff.	Standard error	p-value	Coeff.	Standard error	p-value
Effets fixes									
Constante	4.3487	(0.013858)	<0.001	4.4277	(0.014326)	<0.001	1.9593	(0.075090)	<0.001
Temps				-0.0789	(0.002846)	<0.001	-0.1492	(0.018031)	<0.001
Niveau initial (score EC 6P [Maths])							0.0868	(0.002359)	<0.001
Fille							-0.0001	(0.023458)	0.995
<i>Catégorie socio-professionnelle (réf. CSP moyenne) :</i>									
CSP favorisée							0.2104	(0.029666)	<0.001
CSP défavorisée							-0.1049	(0.026102)	<0.001
Allophone							0.0065	(0.024262)	0.790
<i>Regroupement de 7^e CO (réf. Regroupement A) :</i>									
Regroupement B							0.2408	(0.040439)	<0.001
Regroupement C							0.7113	(0.099199)	<0.001
Classes hétérogènes							0.3304	(0.031095)	<0.001
Temps x Fille							0.0228	(0.005641)	<0.001
Temps x Niveau initial							0.0024	(0.000578)	<0.001
Temps x Regroupement B							0.0160	(0.010022)	0.109
Temps x Regroupement C							-0.0491	(0.022564)	0.030
Temps x Classes hétérogènes							-0.0438	(0.008221)	<0.001
Effets aléatoires									
<i>Niveau 2 (interindividuel [élèves])</i>									
Variance des constantes σ_{u0}^2	0.69242		<0.001	0.74072		<0.001	0.47970		<0.001
Variance des pentes Temps σ_{u1}^2				0.02063		<0.001	0.02000		<0.001
<i>Niveau 1 (intra-individuel [mesures])</i>									
Variance σ_e^2	0.04517			0.01833			0.01833		
Déviance	11316.9			9008.1			7185.6		

Source : SRED.

Figure 23. Modélisation des courbes de croissance en 7^e CO pour les résultats trimestriels de français

(Volée 2007 – Ensemble des regroupements A, B, C et classes hétérogènes)

Paramètres	Modèle 1 « vide »			Modèle 2 inconditionnel de croissance			Modèle 3 conditionnel de croissance		
	Coeff.	Standard error	p-value	Coeff.	Standard error	p-value	Coeff.	Standard error	p-value
Effets fixes									
Constante	4.3981	(0.011440)	<0.001	4.4283	(0.011985)	<0.001	1.0166	(0.075907)	<0.001
Temps				-0.0302	(0.002428)	<0.001	-0.0282	(0.003788)	<0.001
Niveau initial (score EC 6P [Fr1+Fr2])							0.0319	(0.000695)	<0.001
Fille							0.1534	(0.018702)	<0.001
Catégorie socio-professionnelle (réf. CSP moyenne) :									
CSP favorisée							0.0837	(0.022949)	<0.001
CSP défavorisée							-0.0333	(0.019905)	0.094
Allophone							-0.0236	(0.018598)	0.204
Regroupement de 7 ^e CO (réf. Regroupement A) :									
Regroupement B							0.4935	(0.035731)	<0.001
Regroupement C							0.7555	(0.079331)	<0.001
Classes hétérogènes							0.2949	(0.025375)	<0.001
Temps x Fille							0.0137	(0.004818)	0.004
Temps x Regroupement B							-0.0441	(0.007747)	<0.001
Temps x Regroupement C							-0.0556	(0.018368)	0.002
Temps x Classes hétérogènes							-0.0084	(0.007024)	0.232
Effets aléatoires									
Niveau 2 (interindividuel [élèves])									
Variance des constantes σ_{u0}^2	0.47203		<0.001	0.51758		<0.001	0.30694		<0.001
Variance des pentes Temps σ_{u1}^2				0.01505		<0.001	0.01475		<0.001
Niveau 1 (intra-individuel [mesures])									
Variance σ_e^2	0.02918			0.01325			0.01325		
Déviance	6680.5			5188.7			2813.9		

Source : SRED.

Figure 24. Modélisation des courbes de croissance en 7^e CO pour les résultats trimestriels d'histoire

(Volée 2007 – Ensemble des regroupements A, B, C et classes hétérogènes)

Paramètres	Modèle 1 « vide »			Modèle 2 inconditionnel de croissance			Modèle 3 conditionnel de croissance		
	Coeff.	Standard error	p-value	Coeff.	Standard error	p-value	Coeff.	Standard error	p-value
Effets fixes									
Constante	4.5465	(0.012174)	<0.001	4.5538	(0.013109)	<0.001	4.5373	(0.020843)	<0.001
Temps				-0.0074	(0.003536)	0.037	-0.0219	(0.005483)	<0.001
Niveau scolaire initial (score Z moy. EC 6P [Fr1, Fr2, Maths])							0.4720	(0.012633)	<0.001
Fille							-0.0190	(0.021400)	0.374
Catégorie socio-professionnelle (réf. CSP moyenne) :									
CSP favorisée							0.0652	(0.025811)	0.012
CSP défavorisée							-0.1198	(0.023138)	<0.001
Allophone							-0.0559	(0.021543)	0.009
Regroupement de 7 ^e CO (réf. Regroupement A) :									
Regroupement B							0.2593	(0.038773)	<0.001
Regroupement C							0.4983	(0.089398)	<0.001
Classes hétérogènes							0.1851	(0.028851)	<0.001
Temps x Fille							0.0419	(0.007020)	<0.001
Temps x Regroupement B							-0.0278	(0.011399)	0.015
Temps x Regroupement C							-0.0798	(0.029929)	0.008
Temps x Classes hétérogènes							-0.0048	(0.009052)	0.592
Effets aléatoires									
Niveau 2 (interindividuel [élèves])									
Variance des constantes σ_{u0}^2	0.5250		<0.001	0.60827		<0.001	0.39908		<0.001
Variance des pentes Temps σ_{u1}^2				0.03181		<0.001	0.03119		<0.001
Niveau 1 (intra-individuel [mesures])									
Variance σ_e^2	0.0599			0.02819			0.02819		
Déviance	12421.04			11156.66			9384.38		

Source : SRED.

D'une manière générale, en guise de conclusion, relevons encore une fois la très grande similitude des résultats des analyses de croissance entre les volées 2007 et 2012 à tous points de vue : décomposition de la variance, sens et significativité des effets des différentes variables ou encore pouvoir explicatif des modèles (**Figure 25**). Ces deux volées sont pourtant séparées par une réforme du CO. Les résultats qui précèdent sont-ils le signe de l'existence, au sein du processus d'évaluation des élèves par les enseignants, d'un certain nombre d'éléments invariants totalement indépendants de l'organisation du système de formation ?

Figure 25. Comparaison du pouvoir explicatif des modèles multiniveaux de croissance pour les volées 2007 et 2012

	Volée de 7 ^e CO de 2007			Volée de 9 ^e CO de 2012		
	Part de variance expliquée (en %)			Part de variance expliquée (en %)		
	Français	Mathématiques	Histoire	Français	Mathématiques	Histoire
Niveau 1 (mesures)	54.6%	59.4%	53.0%	53.9%	59.0%	53.0%
Niveau 2 (élèves)	35.0%	30.7%	24.0%	37.8%	41.7%	22.2%

7. Synthèse, discussion et perspectives

Avant de reprendre les principaux résultats et d'essayer de les expliquer, nous rappellerons brièvement quelques principes généraux liés à la notation fondés sur la littérature scientifique, étant donné le rôle important que jouent l'évaluation et la notation à l'école dans les processus d'orientation des élèves (cf. *Introduction*).

De manière générale, l'attribution de notes, phénomène complexe, peut être sujette à un certain nombre de biais que l'on peut regrouper en trois catégories : ceux relevant du système scolaire (classe, établissements), ceux relevant des enseignants (valeurs, normes et arrangements internes ou externes) et enfin, ceux liés aux élèves (caractéristiques individuelles, évaluations antérieures, effet Rosenthal ou prophéties autoréalisatrices, etc.) (Leclercq, Nicaise et Demeuse, 2004). La notation n'est donc pas liée directement aux performances des élèves et ne résulte pas de l'application automatique d'un barème mais plutôt d'un processus de fabrication (Merle, 2007). Non seulement l'enseignant note le travail des élèves, mais il le fait en tenant compte de la classe dans laquelle ils sont scolarisés. Par ailleurs, il doit également préserver des relations pédagogiques. La référence au groupe-classe, aux objectifs à atteindre ainsi que les croyances des enseignants liées à la justice sont des éléments à prendre en compte. L'enseignant s'adapte à son public d'élèves, notamment au niveau du contenu pédagogique des cours. En outre, un certain nombre d'études ont montré qu'à performances identiques, plusieurs variables qui caractérisent les élèves influent significativement sur le jugement des enseignants et notamment l'origine sociale et le retard scolaire. Le genre, par contre, n'a pas un effet univoque puisque ce dernier dépend de la discipline considérée.

Distributions différenciées entre R1, R2 et R3 (mathématiques)

Il est étonnant de constater les grandes similitudes entre les distributions des moyennes trimestrielles de R1 et de R2, qui varient entre 4.2 et 4.1 entre le premier et le troisième trimestre, alors que celles de R3 varient de 4.8 à 4.7.

De manière générale, les enseignants, pour des raisons de crédibilité, excluent les extrêmes pour les moyennes trimestrielles (Merle, 2007). Pour Bressoux et Pansu (2003), les notes ne peuvent être ni trop hautes pour maintenir un certain niveau d'exigence, ni trop basses pour rester crédibles. De même, leur dispersion ne peut être ni trop grande (ce qui aurait pour effet de nécessiter des contrats didactiques différents), ni trop faible (pour des besoins de discrimination). Cela pourrait expliquer une utilisation relativement réduite de l'échelle des notes mais pas forcément les différences entre les distributions de R1 et R2 et celle de R3.

Une autre explication pourrait résider dans le fait que si l'on donne aux élèves de R1 et R2 des notes trop élevées, ils peuvent prétendre à changer de regroupement (certains l'ont d'ailleurs fait et ont été supprimés des analyses, les moyennes ne pouvant être comparées d'un trimestre à l'autre). Pour les enseignants, il est nécessaire de trouver un équilibre : ne pas décourager les élèves en leur donnant des notes trop basses, ni les surévaluer et leur donner de faux espoirs de transfert. De plus, on peut se demander si les objectifs visés et les niveaux d'exigence en mathématiques sont si différents d'un regroupement à l'autre en 9^e. De ce fait, les enseignants pourraient estimer que les connaissances des élèves de R1 et R2 sont lacunaires par rapport aux attentes de 9^e en mathématiques, dont les résultats en R3 seraient la référence privilégiée.

Ces différences entre les moyennes de R3 et celles des deux autres regroupements pourraient s'expliquer par des éléments liés aux représentations des enseignants concernant le niveau des acquis potentiels de leurs élèves. Les élèves qui n'avaient pas les notes en fin de 8P pour entrer en R3 ont un niveau moyen à faible et, de ce fait, on ne peut donc pas leur donner des notes trop élevées. Cela

rejoint les constats de Merle. On peut supposer qu'au début du CO, les programmes des regroupements sont encore assez proches. De plus, environ une moitié des maîtres enseignent dans les différents regroupements. Ils se réfèrent alors non seulement au niveau moyen de la classe mais également à leurs représentations des différentes classes ou regroupements.

Cette tendance à donner des notes ou des moyennes plus élevées en R3 rejoint les constats effectués par Bain (1979) dans sa thèse sur la volée de 7^e de 1967 (à savoir une tendance à plus de sévérité des maîtres en G qu'en LS). Pour lui, cela s'expliquerait par une norme de l'excellence gymnasiale mieux définie et mieux intégrée par les enseignants (secondaires) que celles d'autres sections, notamment du fait de leurs propres expériences scolaires.

Mêmes types d'évolution d'un trimestre à l'autre dans les trois regroupements

Si les moyennes sont systématiquement plus basses en R1 et R2 qu'en R3, elles suivent les mêmes tendances, à savoir une légère supériorité au premier trimestre, certainement parce que l'enseignant doit progressivement ajuster son barème à son enseignement et à sa classe pour encourager les élèves, alors qu'au troisième trimestre, c'est l'orientation qui prime, et on doit mesurer ce qui est acquis afin de déterminer l'orientation pour l'année suivante. En mathématiques en tout cas, cette légère baisse s'explique probablement complètement par l'intégration de l'épreuve commune (EC) dans la moyenne du troisième trimestre : par exemple en R3, la moyenne des notes du troisième trimestre (y compris l'EC) est de 4.73, alors que la moyenne des notes de l'EC est de 4.62 ; la note de l'EC, qui entre environ pour un tiers dans la note du troisième trimestre, baisse légèrement la moyenne de ce trimestre. On observe le même phénomène dans les deux autres regroupements (respectivement en R1, 4.08 pour la moyenne annuelle et 3.6 pour l'EC et en R2, 4.08 pour la moyenne annuelle et 3.8 pour l'EC). Si la tendance est la même dans les trois regroupements, les écarts sont toutefois plus importants en R1 puisque la différence entre la moyenne annuelle et la note moyenne à l'EC est de près d'un demi-point. On peut faire le même type de constat pour le français.

Une des explications de ces écarts tient sans doute au fait que l'EC évalue les connaissances et les compétences du programme de l'année, tandis que les moyennes trimestrielles portent sur plusieurs travaux, dont certains à fonction formative, et sur des portions limitées du programme.

Il faut toutefois rappeler que dans cette étude, on ne compare pas des performances à un même contrôle passé trois fois, mais des moyennes trimestrielles qui portent sur des contenus différents. Ce que l'on compare donc, c'est l'adéquation des acquis des élèves aux exigences liées au plan d'études (c'est-à-dire ce qui devrait être acquis ou considéré comme acquis aux différents moments de l'année).

Effet du genre et de l'origine sociale

De manière assez transversale, on relève que l'origine sociale et, dans une moindre mesure, le genre, ont dans quasiment tous les modèles mis en œuvre un effet propre statistiquement significatif. Une origine sociale modeste est associée à une note moins élevée, et les filles tendent à avoir des notes qui baissent moins que les garçons au cours de l'année scolaire. Bien que ces deux caractéristiques soient déjà prises en compte dans le niveau initial et le regroupement (les filles et les élèves issus des milieux les plus favorisés obtiennent de meilleurs résultats scolaires), elles conservent encore un effet propre, certes plutôt modeste, dans quasiment tous les modèles d'analyse. Cet effet est une illustration de la mécanique de l'inégalité des chances selon l'origine sociale ou le genre. Il est rarement fort, mais continu et cumulatif à chaque étape de la formation, et produit sur l'ensemble des parcours de formation des différenciations finalement assez fortes.

Effet « positif » d’être en R1 ou R2 à niveau initial comparable ?

Les élèves de R1 et ceux de R2 auraient une évolution de leurs moyennes différente de celle de leurs camarades de R3, « toutes choses égales par ailleurs » (notamment le niveau initial en fin de primaire et la catégorie socioprofessionnelle). En fait, bien que pénalisé par son orientation en R1 (cf. supra), un élève, compte tenu de son niveau initial, obtient une moyenne trimestrielle de mathématiques en 9^e plus élevée lorsqu’il se trouve en R1 et en R2 plutôt qu’en R3. La situation est en apparence « plus favorable » pour certains élèves au niveau de la note obtenue dans le contexte du regroupement concerné, mais pas au niveau des acquis des élèves. Il faut toutefois rappeler que, de facto, les élèves n’avaient pas le même niveau initial ni les mêmes catégories socioprofessionnelles. La neutralisation de ces deux variables dans l’équation de régression reste une opération difficile à saisir. On peut supposer que les enseignants de R1, voire ceux de R2, apprécient de manière plus favorable les résultats des élèves dans leur notation par rapport aux R3, peut-être parce qu’ils tiennent mieux compte des difficultés des élèves et adoptent un barème visant à les encourager plutôt qu’à évaluer leurs résultats au plus juste. Par ailleurs, on peut également faire l’hypothèse que la probabilité de progresser est plus grande quand on part de notes basses que de notes élevées.

Effet du temps

Quelles que soient les volées et les disciplines, on observe que, d’une part, les notes sont plus élevées dans les regroupements plus exigeants et que, d’autre part, la diminution moyenne des résultats trimestriels au cours de l’année scolaire est moins prononcée dans les regroupements les moins exigeants.

« Toutes choses égales par ailleurs », être scolarisé dans un regroupement moins exigeant permet plus souvent à l’élève de conserver une note stable tout au long de l’année. Ceci montre une forme d’adaptation de la notation aux exigences plus limitées de ces regroupements, mais dans des limites qui restent néanmoins en deçà de la note qui permettrait aux élèves d’envisager une orientation promotionnelle dans un regroupement plus exigeant. Soulignons toutefois que certains élèves de ces regroupements les ont quittés (généralement au premier trimestre) en raison de leurs bonnes notes (moyenne supérieure à 4.8) mais qu’ils n’ont pu être pris en compte dans les analyses faute de recensement exhaustif de leurs résultats scolaires.

Cette double différence, à savoir des notes plus élevées dans les regroupements plus exigeants et une diminution moyenne des résultats trimestriels au cours de l’année moins prononcée dans les regroupements aux exigences moindres, peut, par hypothèse, résulter d’un effet d’adaptation partielle de la notation au niveau des élèves et au niveau d’exigence du regroupement. On peut notamment y voir le résultat d’une transaction interne dans le cadre de la « négociation didactique » (Chevallard, 1986), modulé par des contraintes externes à la classe que sont les normes de transferts dans le regroupement plus exigeant (transaction externe). C’est vraisemblablement pourquoi le processus d’évaluation, sujet à ces contraintes multiples (Leclercq, Nicaise et Demeuse, 2004 ; Merle, 1996), semble se dérouler dans les mêmes termes dans les deux modèles de différenciation (avant ou après la réforme des structures) et pour toutes les disciplines, qu’elles aient ou non un statut de discipline principale.

Des résultats similaires en mathématiques, français et histoire

L’analyse de trois disciplines au statut différent (deux sont des « branches principales ») ne montre pas de différence de fonctionnement. Par ailleurs, selon les disciplines, les moyennes trimestrielles reposent sur un nombre plus ou moins important de travaux (par exemple en histoire, il y a moins d’heures hebdomadaires qu’en français ou en mathématiques et moins de travaux).

On peut faire l’hypothèse que les mécanismes décrits ci-dessus sont inhérents à tout acte de notation dans le contexte scolaire, indépendamment de la discipline considérée. D’ailleurs, quand on observe

l'évolution simultanée des moyennes trimestrielles de mathématiques, français et histoire, on constate que ces moyennes ont une forte tendance à covarier et ceci de façon encore plus marquée en fin d'année, ce qui pourrait peut-être s'expliquer par un affinement de l'évaluation et une confirmation des compétences au bout d'une année dans la même classe au CO. Une autre explication réside dans la présence d'un « socle de compétences scolaires » reposant sur des compétences verbales et de raisonnement qui sont nécessaires à la réussite dans toutes les disciplines (cf. la notion de *compétences transversales*, Rey, 1996).

Comparaison des deux systèmes (volées 2007 et 2012)

Les résultats obtenus en 2007 et 2012 sont globalement similaires. Ainsi, on observe la même décomposition de la variance (et ce pour les trois disciplines) pour les deux volées : au niveau 1 (mesures), 5% pour les mathématiques, 6% pour le français, 11% pour l'histoire ; et le reste au niveau 2 (élèves), respectivement 95%, 94% et 89%.

Une autre similitude est l'existence de profils d'évolution. L'importance des caractéristiques individuelles est du même ordre de grandeur pour les deux volées et pour les trois branches ; elles jouent donc le même rôle quels que soient l'organisation scolaire et le domaine d'enseignement-apprentissage.

De même, au fil de l'année, on observe une baisse des moyennes en mathématiques, français et histoire.

La seule différence entre les résultats observés en 2007 et en 2012 réside dans les moyennes, qui sont systématiquement plus élevées en 2012 qu'en 2007. On peut faire les hypothèses suivantes pour expliquer ce phénomène : d'une part, les normes de passage en 1^{re} année du CO ainsi que le seuil de suffisance ont changé, augmentant en principe les exigences, et les enseignants ont dû adapter en conséquence leur notation ; d'autre part, les regroupements ne se recouvrent pas quand on compare les deux volées : le regroupement A de 2007 se composait d'environ 75% des élèves, ce qui n'est pas le cas pour le regroupement R3 de 2012, qui scolarise une proportion moins importante d'élèves (environ deux tiers) ; le niveau moyen est donc a priori plus élevé.

La grande similitude de résultats entre 2007 et 2012 malgré des changements d'organisation du CO et de seuil de suffisance nous interpelle. On peut faire l'hypothèse que ces changements ne portent pas directement sur les pratiques d'évaluation des enseignants mais plutôt sur des critères d'exigences, c'est-à-dire davantage à un niveau structurel que didactique. L'attribution des notes est fonction d'un jugement professionnel « transversal ». Les enseignants se sont donc adaptés aux nouveaux règlements en se basant sur leurs normes et valeurs ainsi que sur leur expérience. Par ailleurs, bon nombre d'entre eux enseignent dans plusieurs regroupements. La comparaison quasi quotidienne entre les trois types de classes tend donc à se faire, en ce qui concerne la notation, au détriment de la filière à niveau scolaire moindre, par effet négatif de contraste. La norme de référence de l'évaluation serait donc pour ces enseignants (mais aussi peut-être pour les autres) les performances de la filière supérieure.

Un autre constat intéressant à relever réside dans l'importance du niveau initial des élèves au-delà de leurs caractéristiques individuelles et notamment de leur milieu d'origine. Les apprentissages au CO dépendent, dans une large mesure, des connaissances acquises précédemment tout en étant encore et toujours influencés par l'origine sociale ou par d'autres caractéristiques sociodémographiques.

S'il est difficile de tirer des enseignements de l'ensemble de ces résultats, il serait intéressant de poursuivre cette investigation et d'observer l'évolution de la 9^e à la 11^e. Il est fort probable qu'on constaterait des différences de moyennes plus importantes d'une année à l'autre qu'à l'intérieur d'un degré. Enfin, compte tenu des nombreux biais liés à l'attribution de notes mis en évidence dans la littérature scientifique, il paraît nécessaire de poursuivre la réflexion sur l'évaluation en développant la formation initiale et continue des enseignants ainsi que les échanges sur ce thème à l'intérieur des établissements.

Références bibliographiques

- Allal, L. & Lafortune, L. (2008). A la recherche du jugement professionnel. In L. Lafortune et L. Allal (Ed.), *Jugement professionnel en évaluation : pratiques enseignantes au Québec et à Genève* (pp. 187-201). Québec : Presses de l'Université du Québec.
- Allal, L. & Mottier-Lopez, L. (2008). Mieux comprendre le jugement professionnel en évaluation : apports et implications de l'étude genevoise. In L. Lafortune et L. Allal (Ed.), *Jugement professionnel en évaluation : pratiques enseignantes au Québec et à Genève* (pp. 223-239). Québec : Presses de l'Université du Québec.
- Bain, D. (1979). *Problèmes de l'orientation scolaire et fonctionnement d'un cycle d'orientation*. Berne : P. Lang.
- Borko, H. & Cadwell, J. (1982). Individual differences in teachers' decision strategies: An investigation of classroom organization and management decisions. *Journal of Educational Psychology*, vol. 74, no 4, p. 598-610.
- Bressoux, P. (1994). Note de synthèse : Les recherches sur les effets-écoles et les effets-maitres. *Revue française de pédagogie*, no 108, 91-137.
- Bressoux, P. (2004). Formalisation et modélisation dans les sciences sociales : une étude de la construction du jugement des enseignants. *Revue française de pédagogie*, 148, 61-74.
- Bressoux, P. (2010). *Modélisation statistique appliquée aux sciences sociales*. Bruxelles : De Boeck Université (2^e éd.).
- Bressoux, P. & Pansu, P. (2003). *Quand les enseignants jugent leurs élèves*. Paris : PUF.
- Crahay, M. (1996). *Peut-on lutter contre l'échec scolaire ?* Bruxelles : De Boeck Université.
- Chevallard, Y. (1986). Vers une analyse didactique des faits d'évaluation, in J.-M. De Ketele (éd.), *L'évaluation : approche descriptive ou prescriptive ?* (pp. 31-67). Bruxelles : De Boeck-Wesmael.
- Dompnier, B., Pansu, P. & Bressoux, P. (2011). L'évaluation scolaire : une activité multidéterminée. In F. Butera, C. Buchs, C. Darnon (dir.), *L'évaluation, une menace ?* Paris : Presses Universitaires de France.
- Dormont, B. (1999). *Introduction à l'économétrie*. Paris : Montchrestien.
- Dumay, X. & Dupriez, V. (2009). *L'efficacité dans l'enseignement. Promesses et zones d'ombre*. Bruxelles : De Boeck Université (1^{re} éd.).
- Grisay, A. (1984). Les mirages de l'évaluation scolaire. Rendements en français, notes et échecs à l'école primaire ? *Revue de la Direction Générale de l'Organisation des Etudes*, 1984, XIX, 5, pp. 29-42.
- Kronig, W. (2012). Über das Eigenleben von Leistungsbewertungen. In S. Fürstenau & M. Gomolla (Hrsg.), *Migration und schulischer Wandel: Leistungsbeurteilung*. Wiesbaden: VS Verlag für Sozialwissenschaften, 51-64.
- Leclercq, D., Nicaise, J., Demeuse, M. (2004). Docimologie critique : des difficultés de noter des copies et d'attribuer des notes aux élèves. In M. Demeuse (éd.), *Introduction aux théories et aux méthodes de la mesure en sciences psychologiques et en sciences de l'éducation*. Liège : Editions de l'Université de Liège.
- Leroy, N. & Bressoux, P. (2008). L'apport des modèles multiniveaux de croissance dans l'évaluation des changements motivationnels à l'entrée au collège. *Actes du 20^e colloque de l'Admées-Europe*, Université de Genève.

- Madon, S., Jussim, L. & Eccles, J. (1997). In search of the powerful self-fulfilling prophecy. *Journal of Personality and Social Psychology*, vol. 72(4), Apr. 1997, 791-809.
- Merle, P. (1996). *L'évaluation des élèves. Enquête sur le jugement professoral*. Paris : Presses Universitaires de France.
- Merle, P. (1998). *Sociologie de l'évaluation scolaire*. Paris : Presses Universitaires de France.
- Merle, P. (2007). *Les notes. Secrets de fabrication*. Paris : PUF.
- Mottier Lopez, L. (2009). Reconnaissance d'une professionnalité : un processus situé d'évaluation interprétative. In A. Jorro (Ed.), *Les défis de la reconnaissance professionnelle : évaluer, valoriser, légitimer* (p. 123-145). Ottawa : Presses Universitaires d'Ottawa.
- Noizet, G. & Caverni, J.-P. (1978). *Psychologie de l'évaluation scolaire*. Paris : Presses Universitaires de France.
- Perrenoud, Ph. (1998). *L'évaluation des élèves*. Bruxelles : De Boeck.
- Rey, B. (1996). *Les compétences transversales en question*. Paris : ESF.
- Snijders, T. & Bosker, R. (1999). *Multilevel analysis: an introduction to basic and advanced multilevel modeling*. London: Sage.

