



Les effets à long terme d'une expérience de réduction de la taille des classes

Pascal Bressoux

Université Pierre-Mendès-France, Grenoble
Laboratoire des Sciences de l'Éducation
Institut Universitaire de France

Journée d'étude : « Analyser l'évolution des résultats des élèves : introduction au modèle multiniveau de croissance »

Groupe Edumétrie, Service de la recherche en éducation DIP, Genève,
Vendredi 20 mars 2015

Une interrogation sur les effets de la taille des classes

Revue des travaux antérieurs

Synthèse établie par Meuret (2001). La baisse des effectifs n'entraîne de gains notables que si :

- 1) Elle est importante
- 2) Elle se produit aux premiers niveaux de la scolarité
- 3) Elle concerne les élèves de bas niveau socio-économique

→ Apparaît comme une solution coûteuse, dont les bénéfices peuvent être faibles ou ne concerner qu'une petite partie de la population scolaire.

Conseils du HCEE à l'adresse du gouvernement français qui reprennent les grandes conclusions du rapport Meuret.

→ Vont fortement inspirer l'expérimentation de réduction des effectifs dont il sera question ici.

Difficultés **d'estimation** car classes réduites accueillent souvent les élèves faibles.

Le projet STAR (*Student/Teacher Achievement Ratio*)

Seule expérience randomisée à grande échelle de réduction de la taille des classes.

A débuté en 1985 dans le Tennessee.

Enfants qui entrent dans les jardins d'enfants (*kindergarten*) assignés aléatoirement dans petites classes (13-17 élèves), dans classes « pleines » (22-26 élèves), ou dans classes « pleines » avec un enseignant assistant à temps plein.

Enseignants assignés aléatoirement dans les classes.

Elèves maintenus dans les mêmes conditions (petite classe, classe pleine, classe pleine avec assistant) pendant une période maximale de 4 années (jusqu'au grade 3) avec un nouvel enseignant assigné aléatoirement chaque année.

Plus de 6 000 élèves de 329 classes dans 79 écoles ont participé à l'expérimentation la première année et près de 12 000 élèves ont participé à un moment donné aux quatre années d'étude.

Quelques résultats du projet STAR

- Les classes à petit effectif ont obtenu **des résultats significativement meilleurs** dans chaque grade et dans chaque discipline.
- Bénéfices plus importants pour **les élèves des minorités** ou les élèves fréquentant les classes des quartiers défavorisés (*inner-city schools*) que pour les élèves blancs scolarisés dans des écoles non urbaines.
- Au 3e grade, **l'écart entre élèves blancs et noirs (i.e. catégories plus ou moins favorisées socialement)** serait réduit de 38 % si tous les élèves étaient dans des classes à petit effectif (Krueger & Whitmore, 2001).

Recherches conduites en France

- En général, pas spécifiées directement pour ce propos (cf. Meuret, 2001);
- Différentes méthodes utilisées:
 - Corrélationnelles (Bressoux, 1996; Leroy-Audouin & Mingat, 1995; Mingat, 1991; Schmitt-Rolland & Thauvel-Richard, 1997; Suchaut, 1996; Verdon & Thauvel-Richard, 1995; etc.)
 - Expérimentales (Bressoux & Lima, 2011; Ecalle, Magnan & Gibert, 2006)
 - Variables instrumentales (Bressoux, Kramarz & Prost, 2009; Piketty, 2004; Piketty & Valdenaire, 2006; Gary-Bobo & Mahjoub, 2013)

La question des effets à long terme

- Peu d'études sur les effets à long terme de la réduction de la taille des classes
 - L'étude STAR indique des effets à long terme pour les élèves ayant fréquenté une classe réduite au moins 3 ans (Finn, Gerber, Achilles & Boyd-Zaharias, 2001).
 - Globalement, pas moins de 70% des effets immédiats qui sont encore observables au grade 8: i.e. 5 ans de rémanence des effets (Nye, Hedges & Konstantopoulos, 1999).

Une expérimentation de réduction de la taille des classes en France

Le Ministère de l'Éducation Nationale français a lancé en 2002-2003 une expérimentation d'envergure visant à réduire la taille des classes de CP (1^{ère} année élémentaire) à 10 élèves dans les zones défavorisées.

Méthode

Participants

- 100 classes expérimentales (8 à 12 élèves par classe dans les faits avec une moyenne égale à 10,45)

- 100 classes témoins (15 à 27 élèves par classe dans les faits avec une moyenne égale à 21,29).

Toutes dans des milieux défavorisés (écoles en zone d'éducation prioritaire)

Les élèves ont été suivis jusqu'à la fin de la 2^e année élémentaire (CE1)

Leurs acquisitions en français-lecture ont été testées 5 fois (avec items d'ancrage).

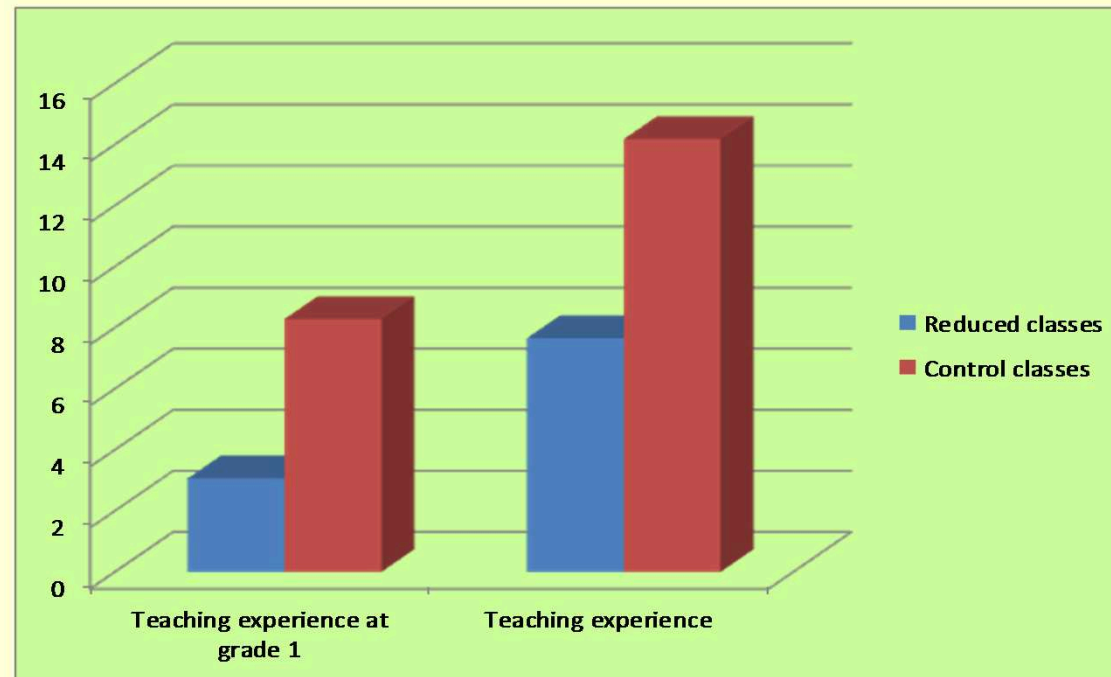
- Début, milieu et fin CP
- Début et fin CE1

(Dans les écoles témoins, ces évaluations n'ont porté que sur 10 élèves choisis aléatoirement.)

Modèles de réponse à l'item ont permis de mettre tous les scores sur une même échelle de mesure.

Tester d'éventuels biais de sélection

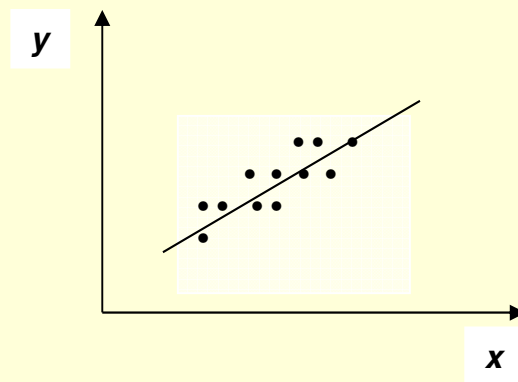
- Caractéristiques initiales des élèves très proches entre les deux groupes
- Enseignants: des différences significatives



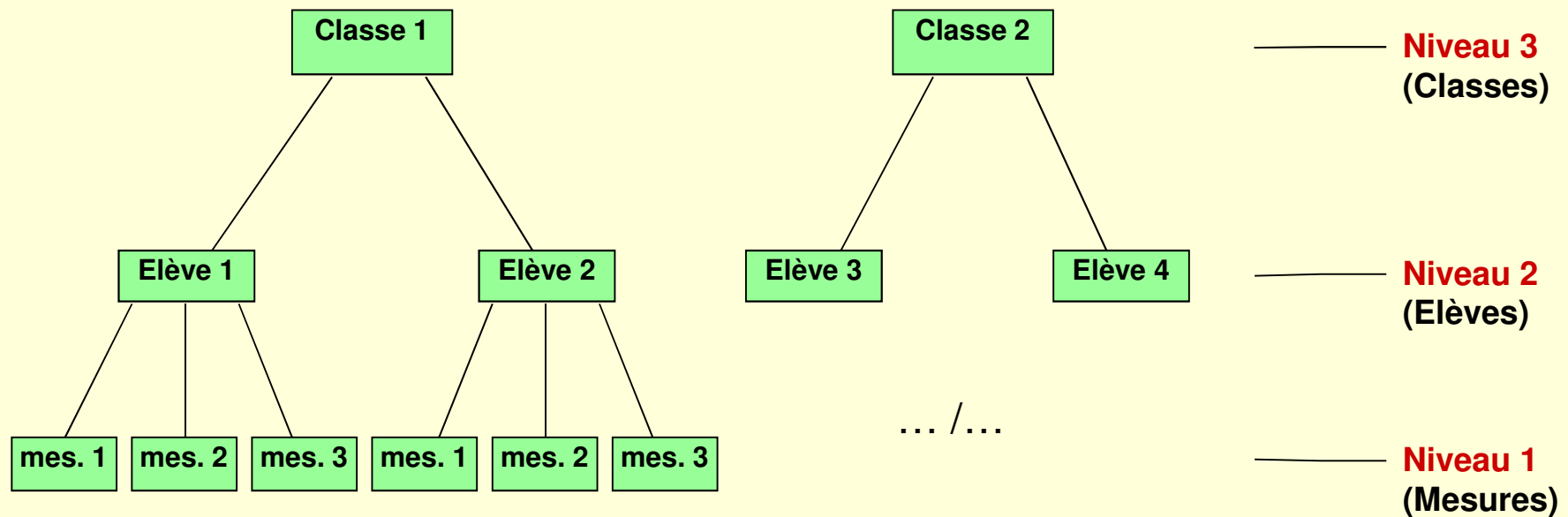
- Donc, besoin de contrôler ces variables (observables)

Spécification du modèle:
Présentation du modèle multiniveau
de croissance

Relation entre les scores initial et final pour un échantillon d'individus



Exemple d'une structure hiérarchisée de croissance



Une mesure du déroulement du temps est nécessaire (âge, durée...)

modèle inconditionnel de croissance

Niveau 1 :
$$Y_{ti} = \pi_{0i} + \pi_{1i} TEMPS_{ti} + e_{ti}$$

Niveau initial moyen

Niveau 2 :
$$\pi_{0i} = \gamma_{00} + u_{0i}$$

$$\pi_{1i} = \gamma_{10} + u_{1i}$$

Rythme de croissance moyen

En intégrant dans une même équation :

$$Y_{ti} = \gamma_{00} + \gamma_{01} Z_i + \gamma_{10} TEMPS_{ti} + (u_{0i} + u_{1i} TEMPS_{ti} + e_{ti})$$

Variance de Y fonction du temps (= gestion de l'hétéroscédasticité des erreurs)

Formalisation du modèle de croissance

Niveau 1 : $Y_{ti} = \pi_{0i} + \pi_{1i} TEMPS_{ti} + \pi_{2i} X_{ti} + e_{ti}$

Niveau 2 : $\pi_{0i} = \gamma_{00} + \gamma_{01} Z_i + u_{0i}$

$$\pi_{1i} = \gamma_{10} + \gamma_{11} Z_i + u_{1i}$$

$$\pi_{2i} = \gamma_{20}$$

Caractéristique qui
varie avec le temps

Caractéristique
interindividuelle
stable dans le temps

En intégrant dans une même équation :

$$Y_{ti} = \gamma_{00} + \gamma_{01} Z_i + \gamma_{10} TEMPS_{ti} + \gamma_{11} Z_i * TEMPS_{ti} + \gamma_{20} X_{ti} (u_{0i} + u_{1i} TEMPS + e_{ti})$$

Rythme de croissance
fonction aussi de Z

Modèle très souple

- Fonctionne pour données non équilibrées (ne nécessite pas le même nombre de mesures par sujet)
- Fonctionne pour des mesures prises à différents moments et dont l'espacement diffère (ne nécessite pas que tous les sujets soient mesurés au même moment).
- Permet de prendre en compte des environnements « macro » pour les individus:
 - classes, écoles, etc. pour les élèves
 - ateliers, usines, etc. pour les ouvriers
 - quartiers, villes, etc. pour les jeunes
 - Circonscription, canton, etc. pour les électeurs
 - Etc.

Attention à la méthode d'estimation

Ce que fait (souvent par défaut) le logiciel :

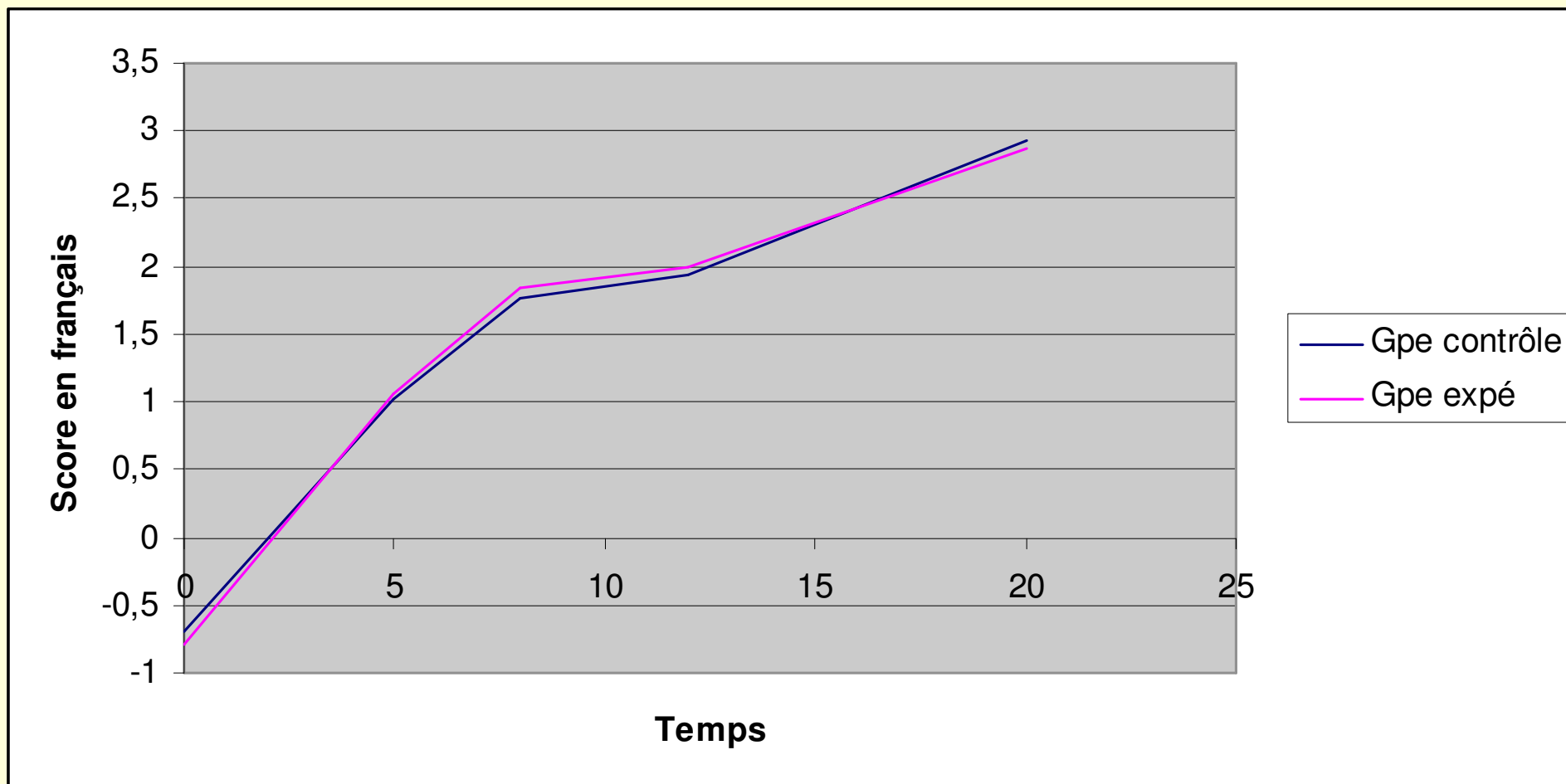
- Maximum de vraisemblance complet (ML ou FML)

Ou

- Maximum de vraisemblance restreint, ou résiduel (RML)

Variable d'analyse : score_francais

temps	Obs	N	Moyenne	N Ecart-type	Minimum	Maximum
0	1163	1163	-0.7470158	1.1387861	-3.6658400	3.5490100
5	1163	1095	1.0424832	0.9507920	-1.4529400	5.8454600
8	1163	999	1.8025968	1.0972123	-2.5008300	5.7569900
12	1163	1013	1.9704838	1.1691808	-0.9668000	5.7569900
20	1163	1163	2.8968559	1.1230747	-0.3306000	6.3903300



Evolution des résultats bruts en fonction du groupe

Spécification du modèle estimant les effets de la réduction de la taille des classes

1163 élèves retenus pour les analyses (i.e. ceux qui étaient présents à la première et à la dernière évaluation).

Structure des données :

Niveau 1 : mesures ($N_t = 5433$)

Niveau 2 : les élèves ($N_i = 1163$)

Niveau 3 : les écoles ($N_j = 69$)

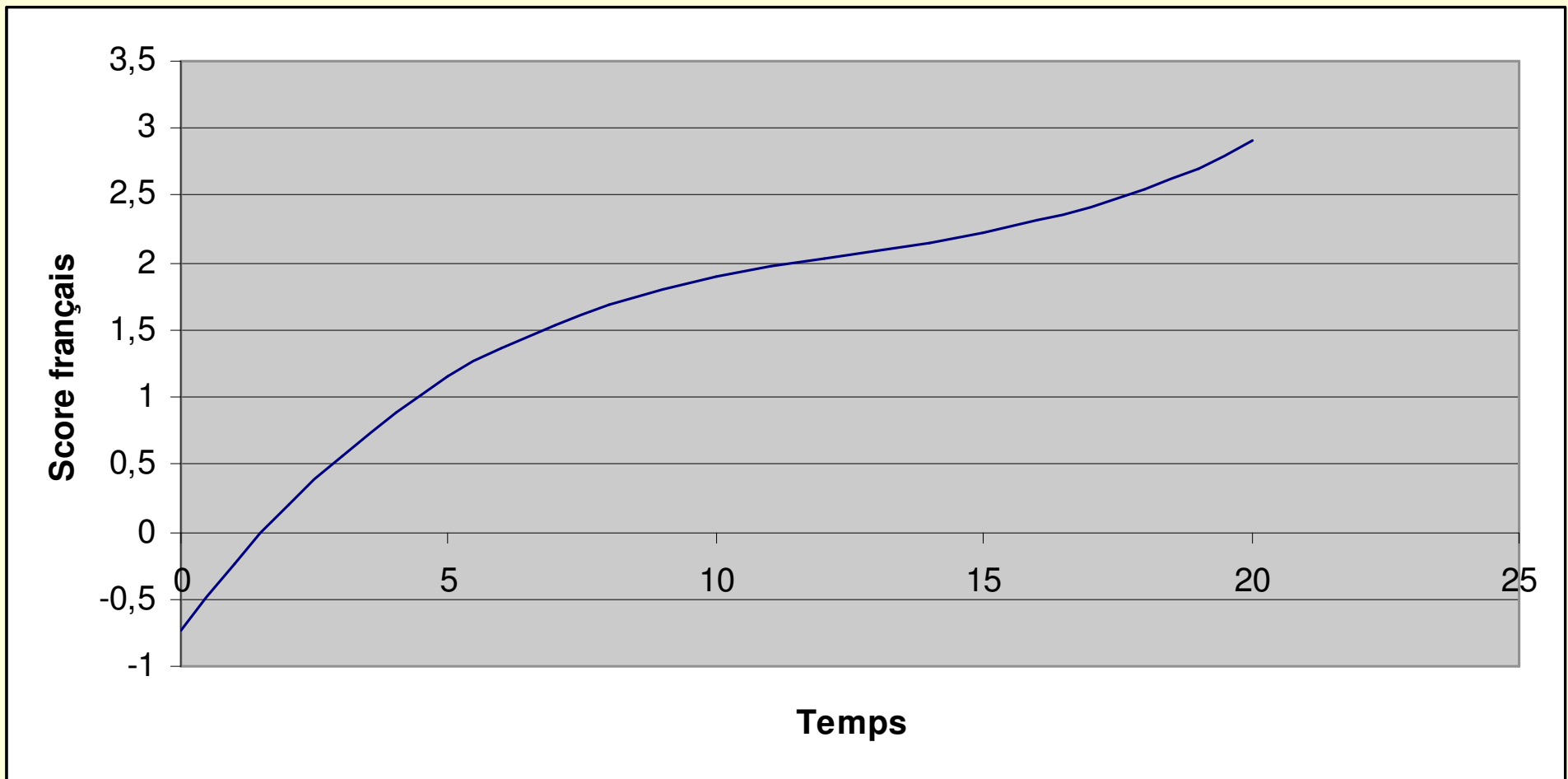
Variables intégrées dans le modèle :

- ▶ Variable *TEMPS* mesurée en nombre de mois (0, 5, 8, 12, 20)
- ▶ Des caractéristiques stables dans le temps (origine sociale, sexe)
- ▶ Des caractéristiques de niveau supérieur (ancienneté enseignant, expérimentation)

Un conseil: travailler finement
la spécification de la forme
temporelle

Quelle spécification adopter?

Une spécification cubique?



Spécification cubique des résultats (sans variables de contrôle)

**Enfin, choix d'une
spécification *piecewise* avec
deux ruptures de pente**

Paramètres	Modèle 1 (inconditionnel = « vide »)	Modèle 2 (inconditionnel de croissance)
Effets fixes		
Constante	0,379 (0,047) ^{***}	-0,147 (0,049) ^{**}
Temps		0,170 (0,002) ^{***}
(Temps – 8)*post-CP		
(Temps – 12)*CE1		
Effets aléatoires		
Variance inter-écoles (niveau 3)		
Variance constantes	0,090 (0,027) ^{***}	0,094 (0,027) ^{***}
Covariance Constantes/temps		
Variance temps	3,22%	6,88%
Variance inter-élèves (niveau 2)		
Variance constantes	0,335 (0,037) ^{***}	0,703 (0,035) ^{***}
Covariance Constantes/temps		
Variance temps	11,90%	49,96%
Variance intra-élèves (niveau 1)	84,79%	43,36%
–2 log L (déviante)	20760,97	14951,46

$N_t = 5433$ (mesures)

$N_j = 1163$ (élèves)

$N_j = 69$ (écoles)

Modèle 1 inconditionnel « vide » :

Variance totale = $0,090 + 0,335 + 2,370 = 2,795$

Variance inter-écoles = $0,090/2,795 = 0,0322$
(3,22 % de la variance totale)

Variance inter-élèves = $0,335/2,795 = 0,1190$
(11,99 % de la variance totale)

Variance intra-élèves = $2,370/2,795 = 0,8479$
(84,79 % de la variance totale)

Modèle 2 inconditionnel de croissance :

Variance totale = $0,094 + 0,703 + 0,610 = 1,407$

Variance inter-écoles = $0,094/1,407 = 0,0688$
(6,88 % de la variance totale)

Variance inter-élèves = $0,703/1,407 = 0,4996$
(49,96 % de la variance totale)

Variance intra-élèves = $0,610/1,407 = 0,4336$
(43,36 % de la variance totale)

Paramètres	Modèle 3	Modèle 4
Effets fixes		
Constante	-0,684 (0,049)***	-0,683 (0,056)***
Temps	0,323 (0,003)***	0,323 (0,004)***
(Temps – 8)*post-CP	-0,304 (0,009)***	-0,305 (0,007)***
(Temps – 12)*CE1	0,097 (0,009)***	0,098 (0,008)***
Effets aléatoires		
Variance inter-écoles (niveau 3)		
Variance constantes	0,090 (0,026)***	0,144 (0,037)***
Covariance Constantes/temps		-0,0052 (0,0017)**
Variance temps	7,39%	0,0005 (0,0001)***
Variance inter-élèves (niveau 2)		
Variance constantes	0,749 (0,036)***	0,791 (0,041)***
Covariance Constantes/temps		-0,0062 (0,0016)***
Variance temps	61,49%	0,0011 (0,0001)***
Variance intra-élèves (niveau 1)	31,12%	0,282 (0,007)***
-2 log L (déviante)	12907,32	12468,15

$N_t = 5433$ (mesures)

$N_j = 1163$ (élèves)

$N_k = 69$ (écoles)

Modèle 3 :

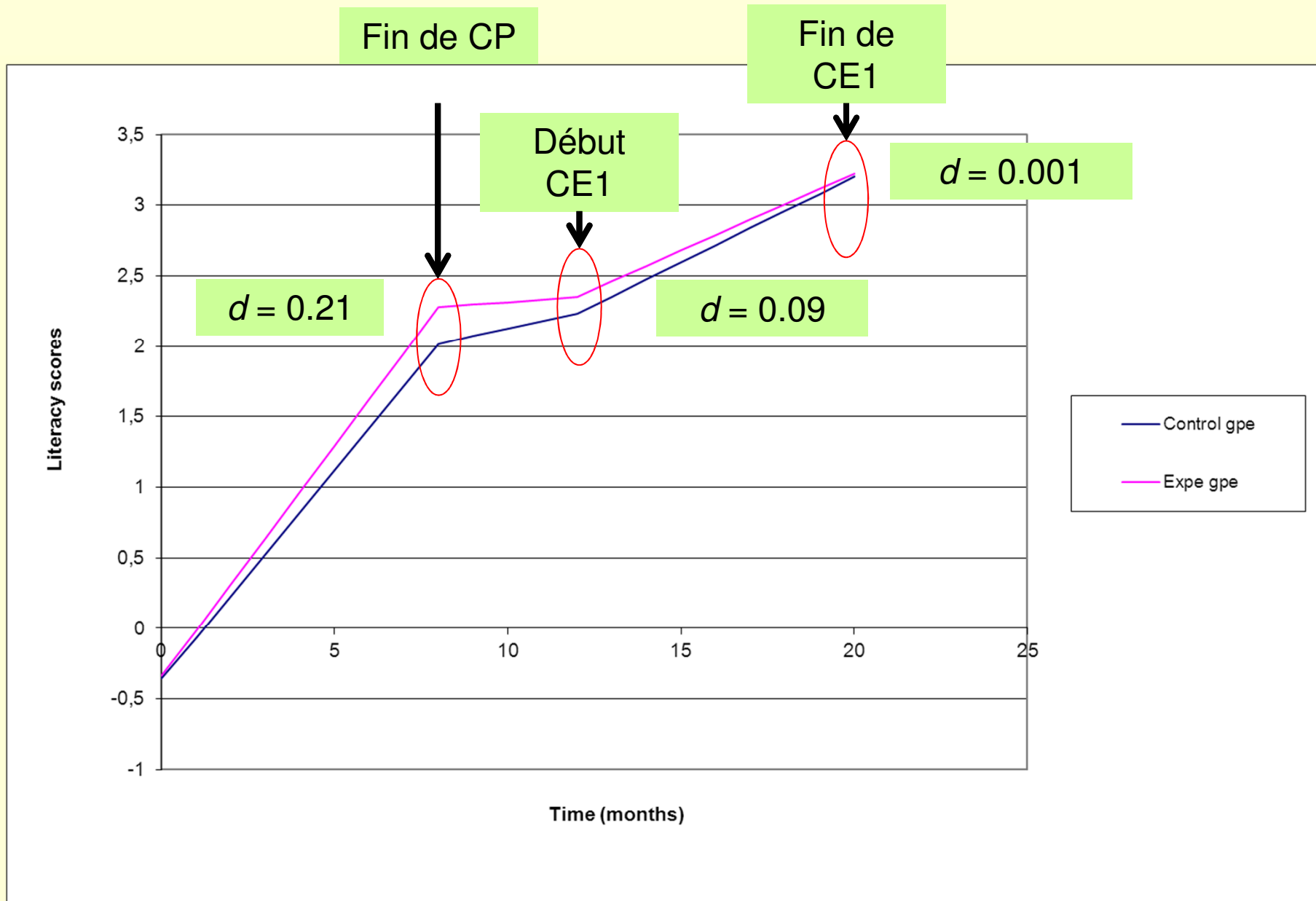
Variance totale = $0,090 + 0,749 + 0,379 = 1,218$

Variance inter-écoles = $0,090/1,218 = 0,0739$
(7,39 % de la variance totale)

Variance inter-élèves = $0,749/1,218 = 0,6149$
(61,49 % de la variance totale)

Variance intra-élèves = $0,379/1,218 = 0,3112$
(31,12 % de la variance totale)

Paramètres	Modèle 5
Effets fixes	
Constante	-0,361 (0,141)*
Temps	0,297 (0,007)***
(Temps – 8)*post-CP	-0,243 (0,014)***
(Temps – 12)*CE1	0,067 (0,014)***
Profession du père (référence = cadre sup.)	
Agriculteur/artisan	-0,399 (0,175)*
Prof. intermédiaire	-0,1575 (0,151)
Employé	-0,319 (0,143)*
Ouvrier	-0,556 (0,127)***
Autre	-0,602 (0,128)***
Fille	0,240 (0,053)***
CP réduit	0,023 (0,097) ns
Ancienneté CP	0,006 (0,006) ns
Temps*CP réduit	0,0295 (0,0075)***
(Temps – 8)*post-CP*CP réduit	-0,0641 (0,0154)***
(Temps – 12)*CE1*CP réduit	0,0225 (0,0158) ns
Temps*Ancienneté CP	0,0019 (0,0005)***
(Temps – 8)*post-CP*Ancienneté CP	-0,0050 (0,0011)***
(Temps – 12)*CE1*Ancienneté CP	0,0034 (0,0011)**
Effets aléatoires	
Variance inter-écoles (niveau 3)	
Variance constantes	0,140 (0,037)***
Covariance Constantes/temps	-0,0054 (0,0017)**
Variance temps	0,0005 (0,0001)***
Variance inter-élèves (niveau 2)	
Variance constantes	0,751 (0,039)***
Covariance Constantes/temps	-0,0065 (0,0015)***
Variance temps	0,0011 (0,0001)***
Variance intra-élèves (niveau 1)	0,277 (0,007)***
-2 log L (déviante)	12353,10



Evolution des scores de français selon le groupe expérimental ou contrôle (la figure est tirée du modèle 5)

Discussion

- Des effets “immédiats” significatifs mais modestes
- Pas d’effets à long terme

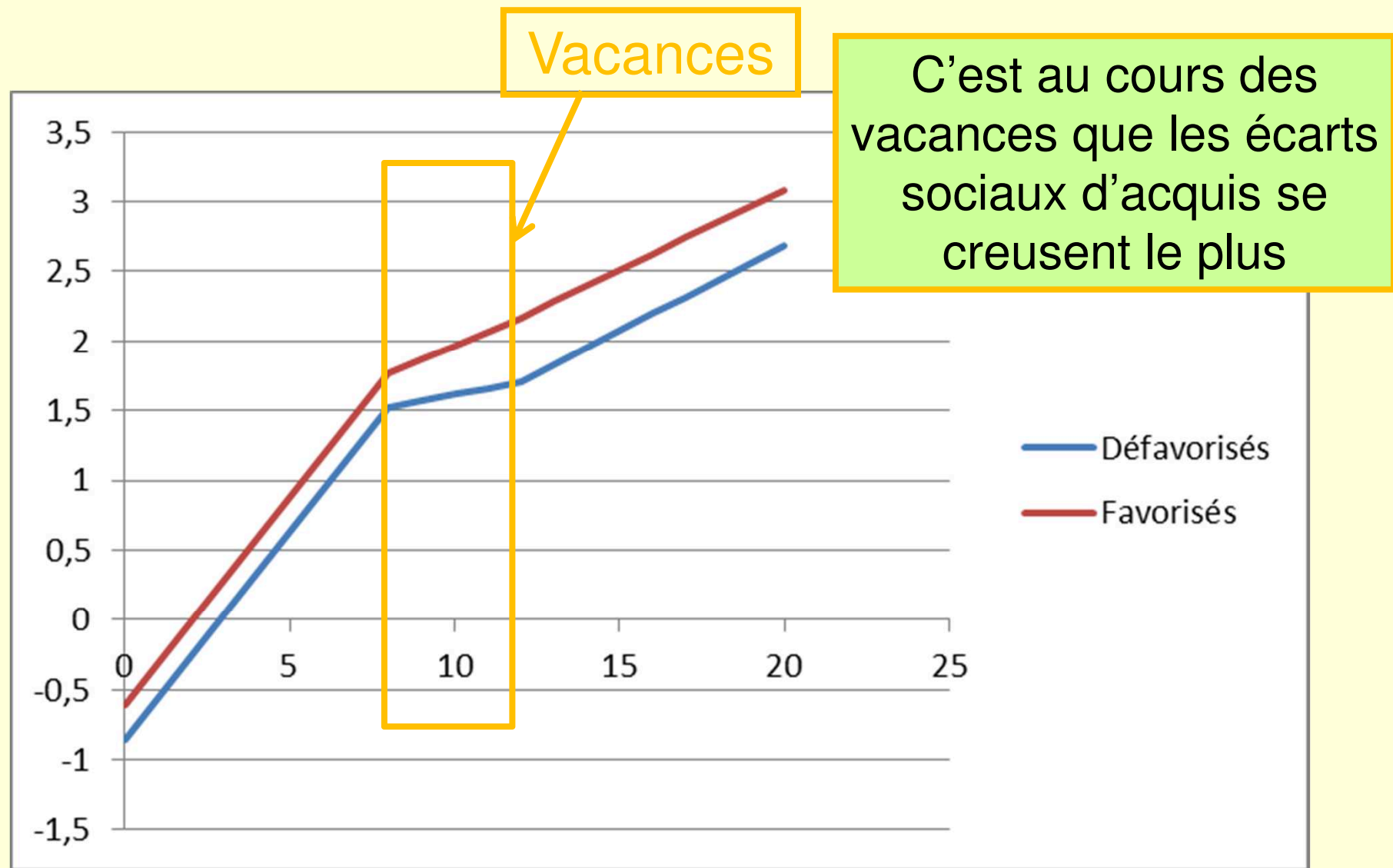
- Pourquoi des effets “immédiats” modestes alors que:
 - Forte réduction (8-12 élèves par classe à comparer avec 13-17 dans STAR)
 - Population d’élèves de zones défavorisées (ZEP)
- Pourquoi pas d’effets à long terme?

- Le protocole expérimental (pas de randomisation):
 - On a contrôlé les variables observables (rappel : contrôler l'expérience des enseignants augmente l'effet expérimental en fin de CP de 60%).
 - Toutes les caractéristiques pertinentes ont-elles été contrôlées?
 - Qu'auraient été les résultats s'il y avait eu randomisation (« contrôle » des variables inobservées)

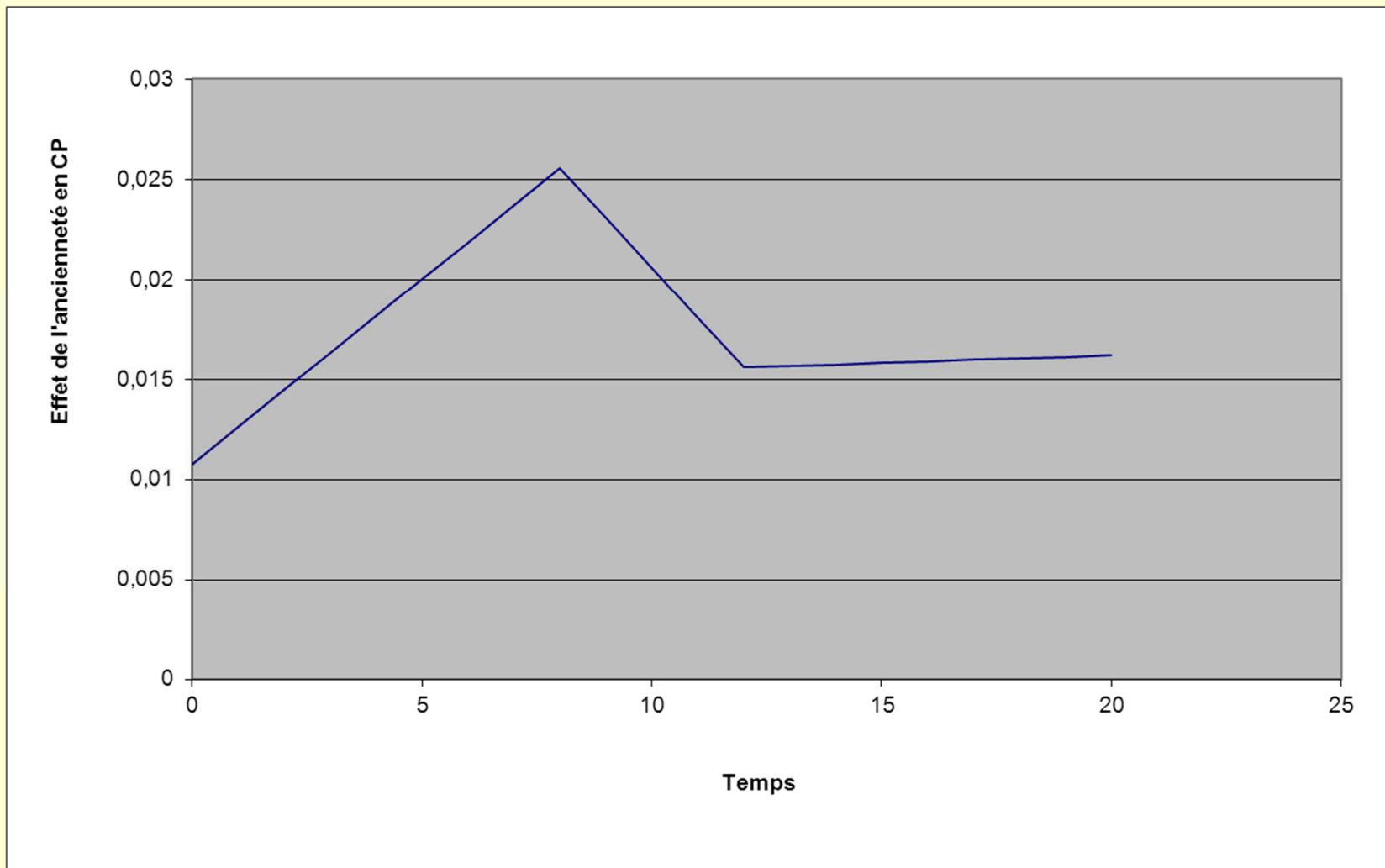
- Concernant spécifiquement les effets à long terme:
 - Les élèves ont fréquenté les classes réduites durant un an (à comparer avec jusqu'à 4 ans dans STAR)
 - => le temps passé dans les classes réduites est-il la clé de l'existence d'effets à long terme?
 - Nos résultats sont cohérents avec ceux de Finn et al. (2001) dans STAR :
“One year (in kindergarten or grade 1) produces early gains but does not produce lasting effects” (p. 174).

Pour explorer d'autres
possibilités de ce genre de
modèles

Par comparaison : la même évolution en fonction de la CSP



Evolution des scores de français selon la catégorie socioprofessionnelle du père




Evolution de l'effet de l'ancienneté sur le score de français

Y a-t-il eu un effet Hawthorne au CP?

- Les enseignants expérimentaux pourraient être plus « motivés » pour démontrer l'efficacité de la réduction des classes

- **Des analyses ont été menées au sein des classes « pleines »** (analyses prétest-post test au CP):
 - Révèlent un effet taille de classe significatif (étendue: 15-27 élèves par classe).
 - L'effet est similaire (et même un peu plus fort) que l'effet sur tout l'échantillon (enlever 1 élève par classe conduit à un gain de 0.025 SD du score de littératie).
- => Probablement pas d'effet Hawthorne

A scenic view of a city, likely Grenoble, featuring a large stone bridge with multiple arches spanning a river. In the background, there are snow-capped mountains under a clear sky. The foreground shows some buildings and a street.

**Merci pour votre
attention**

pascal.bressoux@upmf-grenoble.fr