

Chapitre 10 : Variations sur la même méthode

La méthode que nous avons présentée jusqu'ici est loin d'être figée. A titre d'exemple, il nous a paru utile de présenter succinctement deux variations qui ont été mises en œuvre au sein du PASEC pour des évaluations plus ciblées des systèmes éducatifs de certains des pays membres de la CONFEMEN : les suivis de cohorte et les évaluations thématiques.

A. Suivis de cohorte

Les évaluations que nous avons jusqu'à présent envisagées constituent déjà un court suivi, sur une année scolaire : les mêmes élèves sont testés à deux reprises, au début et à la fin de l'année scolaire. Cela découle de l'objet d'étude : non pas le niveau des élèves, mais bien les apprentissages, ce qui suppose déjà deux mesures au cours du temps.

Mais connaître la trajectoire des élèves sur plus d'une année a de multiples avantages supplémentaires (section 1). Pour cela, un suivi de cohorte, bien que relativement lourd à mettre en œuvre (section 2), s'avère tout à fait irremplaçable. L'analyse des données de cohorte permet alors d'envisager la 'carrière scolaire' des élèves, et en particulier d'étudier les phénomènes de déperdition scolaire (section 3).

1. Les avantages spécifiques d'un suivi de cohorte :

Supposons qu'on veuille évaluer systématiquement l'efficacité de l'enseignement primaire d'un pays dans son ensemble. Envisageons trois démarches possibles :

- Au cours de la même année scolaire, on fait passer des tests de début et de fin d'année à six échantillons d'élèves représentatifs des six niveaux d'enseignement. On collecte à chaque fois l'information habituelle sur les variables contextuelles et les variables de politique scolaire. Les tests, indépendants entre eux, sont fondés sur les programmes de chaque niveau.
- On procède de même, mais en choisissant des tests qui se recoupent partiellement d'un niveau à l'autre.

- On choisit un échantillon représentatif d'élèves de 1^{ère} année du primaire, auxquels on fait passer un test en début et en fin d'année scolaire. A la fin de l'année scolaire suivante, on teste de nouveau les mêmes élèves, du moins ceux qui sont encore à l'école, qu'ils soient passés dans la classe supérieure ou non. On procède de même les quatre années scolaires suivantes, jusqu'à ce que les élèves qui n'ont jamais redoublé atteignent la fin du cycle primaire.

La première approche ne fait que répéter aux six niveaux la stratégie d'évaluation que nous avons étudiée jusqu'à présent. Cela permet de répondre aux questions suivantes :

1. Quel est le **niveau** atteint en moyenne par les élèves qui parviennent à la **fin des différentes années d'études** ?
2. Quelles sont, à chaque niveau du primaire, les facteurs qui expliquent les différences d'apprentissage des élèves ? Répondre à cette question permet de mieux allouer les ressources selon les niveaux d'enseignement : par exemple, les femmes enseignantes ont-elles un avantage comparatif sur les hommes dans les petites classes ? Les maîtres plus diplômés sont-ils spécialement plus performants dans les classes élevées du primaire ? On éclaire ainsi la question de la **variation des facteurs de réussite au cours de la scolarité**.

La seconde approche nous ouvre sur des questions plus dynamiques. Puisque des items communs sont administrés à des élèves de niveaux différents, on va pouvoir constater par exemple que 30% des élèves en fin de CP2 maîtrisent un exercice comportant cinq additions, en réussissant au moins quatre sur cinq, mais qu'ils sont 65% à maîtriser le même exercice en fin de CE1. On peut ainsi dresser un **profil moyen d'acquisition des connaissances et des savoir-faire au fil des niveaux de scolarité**. On répond ainsi à la question :

3. A partir de quel niveau de scolarité est acquise telle ou telle compétence de base ?

Cependant, seule la troisième approche va permettre, par le suivi d'une cohorte d'élèves, de répondre vraiment à la question posée : quelle est l'efficacité du système scolaire primaire pris dans son ensemble ? La première approche, en effet, n'est pas véritablement dynamique : elle considère chaque année isolément, et évalue les compétences des élèves par rapport à une référence liée, de près ou de loin, au programme de chaque année scolaire. La

seconde approche permet certes d'étalonner les progrès moyens des élèves au fil des différents niveaux du primaire. L'inconvénient majeur, cependant, est que le groupe d'élèves qu'on étudie est de plus en plus 'sélecte' au fur et à mesure qu'il se situe dans les niveaux élevés du primaire : au niveau n donné, il comporte un groupe bien particulier d'élèves, ceux qui ont atteint ledit niveau, parfois au terme de plusieurs redoublements, et il exclut un autre groupe, celui de tous ceux qui ont abandonné. Il y a là un **biais de sélection** manifeste lié au phénomène de la **déperdition scolaire**.

Suivre le même groupe d'élèves au cours du temps est le meilleur moyen de prendre en compte ces enjeux afin de répondre vraiment à la question qui nous intéresse :

4. Quelles sont les trajectoires permises par le système d'enseignement primaire ? Comment pourrait-on améliorer ces trajectoires dans les deux dimensions qui définissent l'efficacité d'ensemble d'un système d'enseignement : réduction de la déperdition scolaire (abandons et redoublements) et garantie d'un profil d'acquisition de compétences et d'un profil de sortie satisfaisants ?

2. Spécificité de la mise en œuvre :

Deux aspects requièrent une attention spécifique par rapport aux évaluations réalisées sur une seule année scolaire : le suivi des élèves et le choix d'items d'ancrage au sein des tests.

L'effort de suivi des élèves :

Lorsque les opérations se déroulent sur une seule année scolaire, le suivi des élèves ne pose pas de problème majeur : l'identification de chaque élève est fournie par son nom et sa classe, et le maître peut aider au besoin à retrouver qui a passé le test en début d'année ; les seuls élèves à être perdus de l'échantillon en fin d'année sont les élèves absents du jour, ou ceux qui ont abandonné ou ont été transférés dans une autre école, et sur lesquels le maître peut donner ces informations.

Dans le cas d'un suivi de cohorte sur plusieurs années, retrouver les élèves qui sont encore dans l'école s'avère plus compliqué. Un effort spécifique est indispensable : **la qualité des données issues du suivi de cohorte se joue pour bonne part sur la capacité des enquêteurs à retrouver le maximum**

d'élèves et, lorsque certains ont quitté l'école, à identifier la raison du départ (abandon, transfert vers une autre école,...). Détaillons quelque peu ce point crucial et les méthodes qui peuvent être utiles.

Les élèves, initialement regroupés dans une seule classe, sont à présent dispersés dans plusieurs, selon qu'ils ont ou non redoublé, et selon le nombre de classes qui existent dans l'école pour chaque niveau. La présence d'homonymes au sein de l'école (même nom, même prénom) peut devenir un réel problème. Des précautions doivent être prises, en s'assurant qu'on teste bien le même élève que les années passées. Il est également important de garder pour référence la liste des élèves établie lors de la première année du suivi, de façon à pouvoir retrouver des élèves qui étaient simplement absents le jour du test de l'année précédente. L'idéal évidemment serait de s'appuyer sur un système national d'immatriculation des élèves, qui enlève toute ambiguïté.

Même si de telles précautions sont prises, la cohorte va inévitablement se réduire petit à petit, au fur et à mesure que des élèves quittent l'école. Ces départs devront être intégrés à l'analyse, ce qui nécessite que les causes de l'absence de l'élève soient précisément identifiées : simple absence le jour du test ? abandon ? transfert vers une autre école ? Le directeur semble la personne indiquée pour renseigner l'enquêteur sur ces points. Il peut être utile de préparer la venue de l'enquêteur en envoyant à l'avance au directeur une liste pour recenser les élèves. Voici à titre d'exemple une feuille de suivi des élèves utilisée par le PASEC au Sénégal :

A l'attention du directeur de l'école de
[nom de l'école].

Recensement des élèves présents dans votre école (année 1998-99).

Dans la mesure du possible, il est très important de savoir ce que sont devenus les élèves suivants, **qui avaient été testés en fin de CP lors de l'année scolaire 1995-96.** Merci de compléter ce tableau le plus précisément possible, en cochant **une seule case par élève.**

NOM DE L'ELEVE	L'élève est encore dans mon école				L'élève n'est plus dans mon école			
	En CM1	En CE2	En CE1	Autres (préciser)	Il a été transféré dans une autre école	Il a abandonné l'école	Je ne sais pas ce qu'il est devenu	Autres (préciser)
1								
2								
(...)								
19								
20								

L'utilisation d'items d'ancrage dans les tests :

Le suivi de cohorte doit permettre de :

- Comparer les compétences d'un même élève à différents stades de sa scolarité ;
- Comparer les compétences de deux élèves qui sont dans deux classes de niveaux différents.

Pour que la comparaison soit fiable, il faut que la base de mesure soit strictement la même (mêmes items standardisés, incluant les mêmes consignes de passation et de correction). Cependant, il n'est pas souhaitable, en général, de faire passer en entier le même test à des élèves de CP2 et de CE2 par exemple, car cela obligerait soit à inclure des éléments qui ne figurent pas au programme de CP2, soit à exclure des éléments importants qui figurent au programme de CE2. Pourtant, il est possible et intéressant de comparer la maîtrise de l'addition à ces deux niveaux, par exemple.

Pour ce faire, on conçoit les tests de façon que certains exercices ou items se retrouvent à différents niveaux : on parle d'**exercices ou d'items d'ancrage**. Ces items peuvent porter sur un quart ou un tiers du test ; en fait, lorsque le test a une bonne cohérence interne, les résultats à une fraction du test permettent de bien évaluer les résultats à l'ensemble du test¹.

Concevoir ces items d'ancrage peut sembler délicat : on peut craindre qu'ils soient trop faciles pour les classes élevées par exemple. Ce problème ne doit pas nous arrêter : un bon test est justement fait d'items 'faciles' et 'difficiles' qui permettent de mettre en évidence les compétences variées des élèves. L'autre problème est de parvenir à des questions qui ne soient pas liées étroitement au programme d'un niveau donné, ce qui fausserait la comparaison. C'est un problème qu'il faut savoir résoudre de toute façon, si on veut véritablement tester des 'compétences' des élèves, en vue d'un profil de sortie défini pour l'ensemble de l'enseignement primaire, et pas simplement la maîtrise d'un 'programme'.

¹ Voir la présentation sur la fidélité des tests, au chapitre 3.

3. L'analyse des données : exemple du suivi de cohorte du PASEC en Côte d'Ivoire, sur la question des redoublements :

Limites des données portant sur une seule année pour l'étude du redoublement :

Une façon d'étudier l'efficacité du redoublement est de repérer, dans une évaluation standard portant sur une année scolaire, les redoublants de l'année, et de comparer leurs progrès à ceux des non redoublants. Si les redoublants progressent moins, on tendrait alors à dire que le redoublement est inefficace. Malheureusement, ce n'est pas vraiment concluant, car il se peut très bien que le redoublement soit favorable à l'élève faible, en accélérant ses progrès par rapport à la situation où il ne redoublerait pas, mais sans lui permettre pour autant de progresser au même rythme que l'élève non redoublant moyen. Pour évaluer l'efficacité du redoublement, il faut s'assurer qu'on compare des redoublants et des non redoublants comparables, c'est-à-dire partis du même point de départ : c'est ce que va permettre de faire le suivi de cohorte.

Organisation et lecture préalable des données de cohorte :

Prenons l'exemple de trois années de suivi de cohorte réalisé sur trois années scolaires en Côte d'Ivoire à partir de la deuxième année du primaire (CP2). Comment se présentent les données recueillies ?

Tout d'abord, constatons la réduction progressive des effectifs suivis :

	Début 1995-96	Fin 1995-96	Fin 1996-97	Fin 1997-98
Nombre d'élèves testés	2370	2127	1681	1306

Pour suivre le devenir de chaque élève, il peut être utile d'adopter une procédure de codes à quatre chiffres (un pour chacun des quatre tests). On codera 2 lorsque l'élève est présent le jour du test et en deuxième année, 3 lorsqu'il est présent et en troisième année, 4 lorsqu'il est présent et en quatrième année, 0 lorsqu'il est absent du jour, 8 quand il a été transféré, 9 quand il a abandonné (en supposant que toute cette information soit disponible...). Par exemple, un code d'itinéraire possible est 2204 qui correspond à un élève présent au

second test (fin 1995-96), absent au moment du troisième test (fin 1996-97), présent et au niveau 4 au moment du quatrième test (fin 1996-97). 2238 correspond à un élève présent fin 1995 et en troisième année au moment du troisième test, mais qui a été transféré dans une autre école au moment du quatrième test. Rien que sur trois années de suivi, ce codage permet de dénombrer 32 itinéraires possibles¹ ! Même si l'exercice est laborieux, il prépare utilement l'analyse, lorsqu'il va s'agir d'évaluer les décisions d'orientation ou d'identifier les causes des abandons : il faudra alors savoir comparer aux itinéraires pertinents.

Un exemple d'analyse : l'efficacité du redoublement du CP2 :

Comme toujours, l'analyse va être fondée sur des comparaisons. La question théorique pourrait s'énoncer ainsi : un élève faible progressera-t-il davantage en redoublant ou en ne redoublant pas ? L'expérience comparative idéale consisterait à prendre un élève (ou un groupe d'élèves) en fin de CP2, et, simultanément, de lui faire redoubler le CP2 et de le faire passer en CE1, et de comparer les résultats obtenus au terme de cette année scolaire. Notre maîtrise de l'espace-temps ne nous permet pas encore de réaliser de telles expériences, aussi va-t-on essayer seulement de s'en approcher au mieux. Pour cela, on va identifier deux groupes : un groupe dit de traitement auquel est appliqué le 'traitement' de redoubler le CP2², et un second groupe de contrôle à partir de ceux qui passent en CE1³. Pour que les deux groupes soient vraiment comparables, on restreindra le groupe de contrôle à ceux qui ont le même score que des individus du groupe de traitement, en profitant du fait que les décisions de redoublement comportent une part d'arbitraire : à niveau égal, certains redoublent, d'autres non. A la fin de l'année scolaire, on compare les résultats des deux groupes. Dans la mesure où le point de départ était le même, la différence finale peut être attribuée à l'efficacité ou l'inefficacité du redoublement.

Pour plus de détails, le lecteur peut se référer à l'analyse présentée dans le rapport de synthèse du PASEC dans sa version 1999⁴.

¹ Soit : 2000, 2002, 2003, 2004 ; 2020, 2022, 2023 ; 2030, 2033, 2034 ; 2200, 2202, 2203, 2204 ; 2220, 2222, 2223 ; 2230, 2233, 2234 ; 2888, 2999 ; 2088, 2099 ; 2288, 2299 ; 2208, 2209 ; 2228, 2229 ; 2238, 2239.

² Soit les itinéraires 2220, 2222, 2223, 2228 et 2229.

³ Soit les itinéraires 2230, 2233, 2234, 2238 et 2239.

⁴ Les facteurs de l'efficacité de l'apprentissage dans l'enseignement primaire : les résultats du PASEC sur neuf pays d'Afrique et de l'Océan indien (avril 1999). Voir bibliographie en fin de ce manuel.

B. Evaluations thématiques

1. Principes et objectifs :

Les évaluations présentées jusqu'à présent reposent sur une stratégie que l'on peut qualifier de «généraliste» : il s'agit de réunir dans un échantillon national des conditions d'enseignement aussi diverses que possible, afin de comparer leur efficacité, et ce pour des facteurs aussi divers que le recrutement et la formation des maîtres, les outils pédagogiques, l'environnement socio-culturel local et familial,... Ce type de bilan est utile ; mais les circonstances peuvent conduire à adapter la méthodologie pour une approche plus ciblée : c'est ce que nous appelons les évaluations thématiques.

Dans ces évaluations thématiques, au lieu de chercher l'impact d'un grand nombre de variables sur les apprentissages, l'étude porte sur un ou deux éléments, mais des éléments importants pour lesquels une réforme est mise en œuvre ou est sérieusement considérée. Faire porter l'étude plus précisément sur ces points permet de mettre au point un protocole «sur mesure», ce qui a deux avantages principaux :

- Réduire la taille des échantillons nécessaires, puisqu'on peut s'assurer à l'avance du nombre de classes ou de maîtres représentant les situations qu'on veut comparer ;
- Affiner les questions auxquelles on peut apporter des réponses ; par exemple, il ne s'agira pas seulement d'évaluer l'impact pédagogique des classes à double vacation, mais de mettre à profit l'existence de modes de gestion divers de la double vacation au sein du pays pour comparer leur efficacité respective. Ce gain en précision permet d'entrer dans le détail du “comment faire” d'une réforme. Par exemple : s'il est possible de recruter utilement des maîtres moins diplômés, comment les encadrer ? Si les livres sont bons, quel accompagnement pédagogique permet-il d'en tirer le bénéfice maximal ? Si les classes à double flux permettent de faire face à l'affluence d'élèves, quelles mesures d'accompagnement faut-il prévoir ?

2. Adaptation de la méthodologie aux évaluations thématiques :

Chaque évaluation thématique est spécifique et, comme telle, réclame un protocole spécifique. Il y a cependant une ligne directrice assez générale : on

cherche à comparer des situations qui diffèrent pour un thème d'intérêt (que ce soit la double vacation, le type de recrutement des maîtres,...), mais pour ce thème seulement. Pour s'approcher de ces conditions de l'expérience «toutes choses égales par ailleurs», il est utile de recourir simultanément aux techniques de l'échantillonnage raisonné et de la régression multivariée. Cette ligne directrice peut se détailler en quelques jalons :

- Dans la plupart des cas, les effets recherchés porteront sur les apprentissages des élèves, puisque l'élève est la finalité de tout enseignement. Pour mesurer ces apprentissages, il reste en général nécessaire de faire passer au moins deux tests aux élèves, en début et en fin d'année scolaire.
- Un important travail de recueil d'information doit précéder le choix du protocole d'enquête. L'objectif est double : identifier les enjeux pour la politique éducative sur le thème considéré, et repérer toutes les sources de variation qui peuvent permettre des comparaisons éclairantes (différences régionales dans l'application d'une réforme, décalage dans le temps de certaines mesures qui permet de comparer à une date donnée des pratiques nouvelles et anciennes, donc d'évaluer les réformes en cours,...).
- L'échantillon doit être mis au service des questions posées (**échantillonnage raisonné**). Il est très utile de penser en termes de groupes de comparaison. On aura ainsi au moins un groupe de traitement, qui présente la caractéristique que l'on souhaite évaluer, et un groupe de contrôle, qui sert de référence, et doit être aussi proche que possible du groupe de traitement, sauf pour la caractéristique étudiée. Les techniques usuelles de stratification et de tirage aléatoire peuvent être utilisées au sein de chacun de ces groupes, mais leur objectif n'est plus la **représentativité**, mais bien la **comparabilité** du groupe de contrôle et du groupe de traitement.
- Comme il n'est jamais possible de réunir ces conditions de parfaite comparabilité⁵, il importe de réunir toute une information supplémentaire

⁵ En un sens, mieux le système éducatif est géré, moins il sera possible de trouver un groupe de contrôle et un groupe témoin qui ne diffèrent que par un attribut. Cela voudrait dire, par exemple, qu'une classe à double vacation aurait pu être indifféremment ouverte dans deux écoles, mais l'a été dans une seule. En général, il y a de 'bonnes raisons' qui font que la classe a été ouverte là et pas ailleurs. Ce sont ces 'bonnes raisons' qu'il importe de connaître afin de démêler leurs effets de l'effet de la décision elle-même d'ouvrir la classe à double flux.

sur les variables autres que la variable d'intérêt, mais dont on sait qu'elles jouent en général un rôle important sur les apprentissages. Cela recouvre les variables contextuelles et les variables de politique éducative dont les précédents chapitres ont traité. La **régression multivariée** demeure un outil privilégié pour contrôler l'effet de ces variables afin d'évaluer sans biais les effets de la variable d'intérêt.

3. Un exemple : l'évaluation du programme de recrutement et de formation accélérée de maîtres contractuels en Guinée :

En Guinée (Conakry), le PASEC travaille sur l'évaluation des résultats pédagogiques du programme de Formation Initiale des Maîtres de Guinée (FIMG). Ce programme vise au recrutement de trois fois 2000 maîtres sur 3 ans, avec un niveau académique initial plus élevé (les deux baccalauréats), mais auxquels est dispensée une formation accélérée axée sur la pédagogie et l'animation concrète d'une classe.

Le travail de recueil d'information a permis de mettre en évidence certaines questions pertinentes qui permettent d'affiner la question première de la comparaison de l'efficacité pédagogique de ces nouveaux maîtres par rapport aux maîtres formés de façon classique pendant trois ans dans les Ecoles Normales d'Instituteurs (ENI). On a ainsi repéré l'intérêt de comparer les résultats de ces nouveaux maîtres selon l'ENI dans laquelle ils ont reçu leur formation accélérée. Par ailleurs, le projet insistant fortement sur la supervision que reçoit le nouveau maître dans l'école où il est affecté, il est apparu qu'un paramètre important pouvait être la localisation de cette école d'affectation, son degré d'isolement permettant un contact plus ou moins régulier avec les conseillers chargés de participer à cette supervision. On peut également attendre d'importants effets de génération : la première génération de nouveaux maîtres diffère de la seconde à au moins deux points de vue, celui de la formation reçue, qui était plus rodée la seconde année, et celui du vivier de recrutement, qui avait peut-être commencé à se réduire. Enfin, la dernière génération de maîtres formés en trois ans dans les ENI étant arrivée à sa première affectation la même année que la première génération des nouveaux maîtres, elle est apparue comme un groupe de contrôle assez naturel.

Le travail d'échantillonnage, en cours, devrait donc partir de trois groupes, correspondant aux deux générations de nouveaux maîtres et à la dernière

génération de maîtres formés dans la filière longue. Au sein de ces groupes, des strates peuvent être conduites selon des lignes géographiques (régions, milieu urbain et rural).

Enfin, les questionnaires devraient permettre de préciser à la fois les paramètres dont on pense qu'ils peuvent influencer sur l'efficacité des nouveaux maîtres (importance de l'encadrement, projet de carrière personnel du maître,...) et les variables de contexte dont les effets devront être contrôlés par la régression multivariée.