

Manuel d'échantillonnage, pondération, estimation de la variance & réplifications



pasec
Programme d'analyse des systèmes
éducatifs de la confemen

Équipe du PASEC

Le présent document a été préparé par l'équipe du Programme d'analyse des systèmes éducatifs de la CONFEMEN (PASEC).

- Hilaire Hounkpodoté, Coordonnateur du PASEC
- Bassile Tankeu, Chef de la division « Statistiques, analyse des données et études » du PASEC
- Labass Lamine Diallo, Chef de la division « Conception des instruments, collecte des données et renforcement des capacités » du PASEC
- Ousmane Birba, Conseiller technique du PASEC
- Seydou Garba Hamidou, Conseiller technique du PASEC
- Tchindebé Berkeunbé, Conseiller technique du PASEC
- Valérie Vinuesa, Conseillère technique du PASEC
- Fatimé Zara Boukar, Conseillère technique du PASEC
- Adama Bologo, Chargé de projet du PASEC
- Khady Mbaye Camara, Assistante administrative principale du PASEC
- Penda Seck Fall Diouf, Assistante administrative adjointe du PASEC
- Ulrich Tadajeu Kenfack, Attaché de communication du PASEC

Relecture

Alexis LOYE, Président du comité scientifique du PASEC

Publié en 2026 par le Programme d'analyse des systèmes éducatifs de la CONFEMEN (PASEC)

Complexe Sicap Point E, Immeuble C, 3^e étage, Avenue Cheikh Anta Diop, BP : 3220 Dakar, Sénégal

© PASEC 2026

Ce document est disponible en accès ouvert sous licence Creative Commons Attribution – Partage dans les mêmes conditions 4.0 International – CC BY-SA 4.0, sauf indication contraire.

Toute utilisation, reproduction, adaptation, traduction ou diffusion, totale ou partielle, est autorisée à condition de mentionner la source, de citer correctement la publication et de diffuser toute œuvre dérivée sous une licence équivalente.

Pour citer ce document :

PASEC. 2026. *Manuel d'échantillonnage, pondération, estimation de la variance & répliques*. Dakar : PASEC.

La CONFEMEN encourage l'utilisation et la diffusion des informations contenues dans cette publication à des fins de recherche, de formation, d'élaboration de politiques publiques et de plaidoyer en faveur d'une éducation de qualité.

Conception graphique et mise en page : Suzanne Grossmann

ISBN : 92-9133-226-7

Sommaire

Liste des tableaux et figures	5
Chapitre 1. Plans d'échantillonnage : aspects théoriques	7
1.1. Introduction	8
1.2. Définition de la population	8
1.3. Exclusions	10
1.4. Plans d'échantillonnage et pondération	14
1.5. Taille de l'échantillon	25
1.6. Stratification	27
1.7. Recouvrement des échantillons de 6 ^e année et 2 ^e année du primaire	28
Chapitre 2. Échantillonnage : procédures techniques	30
2.1. Étape1 : la liste des écoles de la population ou <i>Base de sondage</i>	31
2.2. Étape 2 : attribution d'une mesure de taille 'MOS' à chaque école	34
2.3. Étape 3 : calcul des taux d'exclusion d'écoles	35
2.4. Étape 4 : le traitement des petites et très petites écoles	35
2.5. Étape 5 : calcul de la taille de l'échantillon d'écoles pour chaque strate explicite	36
2.6. Étape 6 : sélection des 200 écoles primaires de l'échantillon principal	39
2.7. Étape 6 bis : traitement des très grandes écoles ou « écoles certaines »	41
2.8. Étape 7 : désignation des écoles de remplacement	43
2.9. Étape 8 : sélection des écoles de l'échantillon école de 2 ^e primaire	44
2.10. Étape 9 : attribution d'un identifiant PASEC	45
2.11. Finalisation de la liste des écoles échantillonnées	46
2.12. Vérification du fichier <i>Échantillon</i>	47
Chapitre 3. Échantillonnage des classes et des élèves	49
3.1. Sélectionner la classe et les élèves	50
3.2. Documentation	51
Chapitre 4. Pondération des données	53
4.1. Préparation des fichiers	54
4.2. Calcul du poids initial de l'école	58
4.3. Calcul du taux de participation des écoles	58
4.4. Calcul de l'ajustement pondéral pour la non-réponse des écoles	62
4.5. Pondération classe et pondération élèves	65
4.6. Vérification des pondérations	66
4.7. Calcul du taux d'exclusion et de non-réponse des élèves	67
4.8. Pondération des données de 2 ^e année	67

Chapitre 5. Estimation de la variance d'échantillonnage	69
5.1. Introduction	70
5.2. Variance d'échantillonnage d'échantillons aléatoires et simples	70
5.3. Variance d'échantillonnage d'un échantillon à plusieurs degrés	76
5.4. Les méthodes de réplication pour les échantillons aléatoires et simples	79
5.5. Méthodes de réplication pour les échantillons à deux niveaux	82
5.6. Procédures techniques d'estimation de la variance d'échantillonnage par JK2	86
Bibliographie	89

Liste des tableaux et figures

Tableaux

TABLEAU 1	Moyennes tronquées selon le pourcentage d'exclusion	13
TABLEAU 2	Taille et pondération de 10 individus	17
TABLEAU 3	Estimation pondérée et non pondérée de l'écart-type	17
TABLEAU 4	Probabilités de sélection et poids correspondants pour un échantillonnage à deux niveaux, avec échantillonnage aléatoire et simple au premier niveau école, les écoles étant de tailles égales	20
TABLEAU 5	Probabilités de sélection et poids correspondants pour un échantillonnage à deux niveaux, avec échantillonnage aléatoire et simple au premier niveau école, la taille des écoles variant	20
TABLEAU 6	Probabilités de sélection et poids correspondants pour un échantillonnage à deux niveaux, avec SRS au niveau école – Sélection des quatre plus petites écoles	21
TABLEAU 7	Probabilités de sélection et poids correspondants pour un échantillonnage à deux niveaux, avec SRS au niveau école – Sélection des quatre plus grandes écoles	21
TABLEAU 8	Probabilités de sélection et poids correspondants pour un échantillonnage à deux niveaux, avec PPS au niveau école	22
TABLEAU 9	Sélection d'écoles par PPS selon une procédure systématique	24
TABLEAU 10	Exemple fictif de <i>Base de sondage</i>	35
TABLEAU 11	Différences entre les variances d'échantillonnage selon la taille de l'échantillon par strate	37
TABLEAU 12	<i>Base de sondage</i> fictive : 2 strates de 5 écoles de tailles différentes	38
TABLEAU 13	Échantillonnage d'écoles proportionnel au nombre d'écoles dans la strate	38
TABLEAU 14	Échantillonnage d'écoles proportionnel au nombre d'élèves dans la strate	39
TABLEAU 15	Exemple de tris successifs selon les variables de stratification implicite	40
TABLEAU 16	Sélection PPS systématique de 4 écoles dans une strate de 400 élèves	41
TABLEAU 17	Désignation des écoles de remplacement R1 et R2	44
TABLEAU 18	Exemple d'attribution des identifiants PASEC des écoles	45
TABLEAU 19	Fichier <i>Échantillon</i> (exemple pour un pays qui dispose des tailles pour chaque grade)	47
TABLEAU 20	Profils possibles des quatre variables après fusion des fichiers <i>Échantillon</i> et <i>Liste des écoles</i>	56
TABLEAU 21	Écoles échantillonnées non exclues et écoles de remplacement	56
TABLEAU 22	Écoles de l'échantillon initial et écoles de remplacement participantes	60
TABLEAU 23	Écoles participantes et les écoles de l'échantillon initial non participantes et qui n'ont pas été remplacées	64
TABLEAU 24	Description des 630 échantillons de deux élèves possibles à partir d'une population de 36 élèves	71
TABLEAU 25	Distribution de tous les échantillons possibles dont la moyenne se situe entre 8.32 et 11.68	74
TABLEAU 26	Distribution de la moyenne de tous les échantillons de 4 élèves possibles à partir d'une population de 36 élèves	75
TABLEAU 27	Effet d'échantillonnage et erreur de type I	78
TABLEAU 28	Les échantillons répliqués Jackknife et leur moyenne	80

TABLEAU 29	Valeur des variables X et Y pour un échantillon de 10 élèves	81
TABLEAU 30	Coefficients de régression pour l'échantillon total et les 10 répliques	81
TABLEAU 31	Exemple de matrice d'échantillons Bootstrap.....	82
TABLEAU 32	Réplication Jackknife pour un échantillonnage non stratifié à deux niveaux.....	83
TABLEAU 33	Les répliques Jackknife pour un échantillon stratifié à deux niveaux	85
TABLEAU 34	Réplications selon la méthode JK2.....	86

Figures

FIGURE 1	Population cible, population effective et niveaux d'exclusion.....	11
FIGURE 2	Exclusion des élèves les plus faibles	13
FIGURE 3	Échantillon en grappes à deux niveaux	18
FIGURE 4	Extrait d'une <i>Base de sondage</i> fictive	33
FIGURE 5	Strate de 10 écoles et 400 élèves, échantillonnage de 4 écoles – 1 école certaine	42
FIGURE 6	Illustration du processus d'échantillonnage des écoles.....	43
FIGURE 7	Distribution des résultats des 36 élèves.....	70
FIGURE 8	Distribution d'échantillonnage de la moyenne	73
FIGURE 9	Distribution d'échantillonnage pour l'estimation de la moyenne calculée comme pour un SRS d'écoles et distribution d'échantillonnage attendue pour un échantillon PPS d'écoles	78

Chapitre 1

Plans d'échantillonnage : aspects théoriques



1.1. Introduction

Dans la plupart des études nationales et internationales, les données sont collectées à partir d'un échantillon et non auprès de la population complète. Les avantages de l'échantillonnage par rapport au recensement sont assez évidents (Neyman, 1934) :

- pour être à même de réaliser un recensement, il faut que tous les individus de la population à étudier soient ou puissent être identifiés. Dans certaines circonstances, on ne peut dénombrer tous les individus de la population ;
- l'échantillonnage permet d'économiser de l'argent. De toute évidence, cela coûte moins cher de collecter des informations auprès de 1000 individus plutôt qu'auprès d'une population de 100 000 individus ;
- l'échantillonnage permet encore d'économiser du travail et du temps et en ce sens est plus compatible avec des échéances raisonnables de publication des résultats d'une étude ;
- la plus petite envergure de l'enquête par échantillon facilite le contrôle de la qualité des différents instruments et leur gestion, réduisant ainsi le risque d'erreurs ;
- enfin, au-delà d'un certain seuil, l'augmentation de la taille de l'échantillon ne s'accompagne plus que d'un gain de précision de plus en plus tenu.

Dans toute étude, la sélection d'échantillons de qualité est un élément clé. Seuls les échantillons soigneusement sélectionnés et représentatifs de la population permettent d'obtenir des estimations non biaisées, précises et comparables au niveau international. Pour obtenir un échantillon de qualité, trois questions majeures doivent être considérées :

1. Quelle est la population à étudier ?
2. Quelle est la procédure d'échantillonnage à mettre en œuvre, compte tenu des contraintes pratiques et des questions de recherche ?
3. Quelle taille attribuer à l'échantillon ?

Ces trois questions vont être successivement développées dans les paragraphes suivants.

1.2. Définition de la population

La première étape d'une étude à large échelle consiste à définir la population qui sera évaluée. Cette tâche n'est pas aussi simple et évidente qu'elle y paraît.

Tout d'abord, il faut définir la population cible de l'étude. Dans le cadre des études internationales en sciences de l'éducation, il est d'usage d'opposer les populations cibles définies en fonction d'une année d'études (en anglais, *grade cohort*) ou en fonction d'un âge donné (en anglais, *age cohort*), généralement une période de 12 mois (Adams & Wu, 2003 ; Martin et al., 2003).

Les populations cibles définies en fonction d'une année d'études sont particulièrement intéressantes car, pour un pays donné, il est aisé pour les décideurs politiques et pour les acteurs pédagogiques de bien se représenter cette population cible d'élèves. Ainsi, un ministre ou le corps enseignants, voire même la société civile d'un pays donné n'éprouvent aucune difficulté à se représenter la population des élèves de 6^e primaire (fin primaire) ou de la 9^e année (fin collège). Cette population d'élèves par ailleurs est inscrite dans des établissements généralement primaires ou collèges recensés par les autorités administratives. Dans ces conditions, le responsable national de l'enquête internationale ne rencontre aucune difficulté majeure à définir la population cible et à l'échantillonner.

Par contre, les populations cibles définies en fonction d'une année d'études occasionnent de nombreuses difficultés en termes de comparabilité internationale. En effet, la scolarité obligatoire et/ou la scolarité primaire commence à des moments et selon des règles variables d'un pays à l'autre, voire parfois d'un état ou d'une province à l'autre dans des pays fédérés comme l'Allemagne à titre d'exemple. Ainsi, si l'on s'intéresse, comme l'étude *Progres International Reading Literacy Study* (PIRLS) de l'IEA, à la 4^e année de l'enseignement primaire, on encourt le risque de comparer des élèves âgés de 8 ans en Grande-Bretagne, de 9 ans en Belgique ou en France et de 10 ans en Norvège (von Davier et al., 2023). En effet, la scolarité primaire commence à 5 ans en Grande-Bretagne, à 6 ans en Belgique ou en France, et à 7 ans dans bon nombre de pays du nord de l'Europe. À ce stade de développement, bien que les élèves des différents pays aient bénéficié d'une même

période de scolarité, à savoir 4 ans dans PIRLS, une différence de 2 ans peut avoir un impact substantiel sur le niveau moyen de performance attendu.

Pour limiter de telles variations d'âge moyen d'un pays à l'autre, la population cible dans l'étude PIRLS est définie comme « tous les élèves inscrits dans l'année d'étude qui représente quatre années de scolarité à partir de la première année de scolarité primaire, pour autant qu'au moment de l'administration des tests et questionnaires, l'âge moyen soit de 9.5 ans ». Cette définition, si elle limite partiellement les variations d'âge, oblige certains pays à tester les élèves en cinquième année, altérant ainsi l'uniformité recherchée au niveau du nombre d'années d'études d'un pays à l'autre.

Enfin, comme l'a démontré Gustafsson (2007), les évolutions de performance moyenne observées entre 1995 et 2003 pour l'étude *Trends International Mathematic and Science Study* (TIMSS) de l'IEA corrèle à 0.58 avec l'évolution de l'âge moyen des élèves. Des changements orchestrés dans les politiques éducatives relatives au redoublement ou à la scolarité obligatoire ou un déplacement de la période de collecte des données peuvent engendrer un changement de l'âge moyen des élèves et ainsi altérer la comparabilité des résultats.

Les études qui, comme PISA, adoptent une population cible définie par un âge, évitent ces problèmes de comparaison internationale. En effet, l'âge moyen des élèves est identique d'un pays à l'autre, la variabilité de l'âge des élèves d'un pays à l'autre est aussi identique puisque la population est définie pour une période de 12 mois alors que dans le cadre des échantillons par année d'études, la variabilité de l'âge des élèves sera plus ou moins importante selon les politiques de redoublement en vigueur dans les pays. Par contre, si une population d'âge offre de meilleures garanties en termes de comparabilité internationale, elle occasionne de nombreuses difficultés relatives notamment (i) à la construction de la *Base de sondage*, étant donné que des élèves d'un âge donné peuvent fréquenter différents types d'établissements, (ii) à l'administration des tests au sein des écoles puisque les élèves d'un âge donné au sein d'une école, peuvent fréquenter différents niveaux scolaires, et, pour une année d'études données, différentes classes. Cette répartition des élèves d'un âge donné dans différents types d'établissements, dans différents niveaux scolaires et années d'études engendrent, auprès des responsables politiques ou administratifs ainsi que le corps professoral, de nombreuses difficultés à se représenter correctement cette population scolaire.

Les responsables scientifiques et politiques qui, à la fin des années 80 ou au début des années 90, ont élaboré le programme PASEC, ont opté pour des populations cibles définies en fonction d'une année d'études.

En effet, ces études PASEC avaient pour objectif principal l'identification des facteurs d'efficacité, c'est-à-dire les variables *Écoles* et les variables *enseignants* associées à une meilleure performance des élèves. L'étude des effets enseignants ou des effets *Classes* nécessite une procédure d'échantillonnage qui sélectionne un certain nombre d'élèves au sein d'une ou de plusieurs classes. Il est dès lors possible de cibler le recueil de données seulement auprès des enseignants concernés par la ou les classes sélectionnées. Échantillonner des élèves au sein des écoles à travers les classes nécessiterait d'interroger tous les enseignants concernés par le niveau d'études et, par ailleurs, il ne serait pas aisé de lier les données enseignants avec les données élèves.

Les évaluations groupées du PASEC, mise en place à partir de l'année 2014 ont pour objectif principal :

- d'évaluer le niveau de maîtrise des élèves du curriculum en compréhension de l'écrit et en mathématiques au terme de la scolarité primaire, et d'identifier les facteurs *Élève*, *Enseignant*, *École* et *Pays* associés à ces performances en vue d'améliorer l'efficacité et l'équité des systèmes éducatifs participants ;
- de recenser les difficultés scolaires des élèves en début de scolarité primaire.

Ces évaluations groupées ont donc perpétué la tradition d'évaluer des populations d'élèves d'un niveau d'études donné. Ce choix repose également sur des considérations pratiques puisqu'une population cible définie en fonction de l'âge nécessite de nombreuses communications écrites entre le centre national et les écoles de l'échantillon. En effet, pour garantir la représentativité et le caractère aléatoire des échantillons, chaque école doit transmettre au niveau central la liste des élèves éligibles (c'est-à-dire dans le cas présent d'un âge donné, quelle que soit l'année d'études) afin que ce dernier puisse échantillonner à l'avance les élèves.

Dans le cadre du programme PASEC, les populations cibles (en anglais, *target population*) pour un pays donné se définissent donc comme :

- l'ensemble des élèves inscrits en 6^e année au primaire, quel que soit le type d'écoles (écoles publiques, privées...) ;
- l'ensemble des élèves inscrits en 2^e primaire, quel que soit le type d'écoles ;
- l'ensemble des élèves inscrits en 9^e /10^e année (fin de collège), quel que soit le type d'écoles (écoles publiques, privées...).

Il convient de noter que l'évaluation au niveau du primaire concerne les élèves de 6^e année et ceux de la 2^e année. Par ailleurs, au niveau du collège, l'évaluation est centrée sur les élèves de la 9^e année (fin du collège). C'est donc en fonction de ces populations cibles que le plan d'évaluation sera conçu.

Pour une année d'études donnée, les enquêtes du PASEC collectent des données à trois niveaux :

- au niveau école, via un questionnaire adressé au directeur ;
- au niveau de la classe, via un test et un questionnaire administré à l'ensemble des enseignants de l'école primaire ;
- au collège, un questionnaire est administré uniquement aux enseignants de la langue d'enseignement et des mathématiques de la classe enquêtée. Au niveau élève, via un test et un questionnaire pour les élèves.

Un questionnaire est adressé aux parents de tous les élèves participant à l'étude.

Malgré ces trois niveaux, la population d'intérêt de l'étude est bien constituée des élèves et non des écoles ou des enseignants. Si l'étude collecte des informations au niveau école et enseignant, c'est principalement pour expliquer les variables élève. Ce choix méthodologique aura des implications (i) sur le plan d'échantillonnage en ce sens que celui-ci sera optimisé pour assurer un maximum de précision au niveau élève et (ii) sur les procédures d'analyses (voir plus loin).

Par ailleurs, la collecte de données au niveau des enseignants et des écoles d'un échantillon représentatif d'élèves n'induit pas nécessairement la représentativité de ces deux populations. À titre d'exemple, imaginons une enquête qui recueille des données auprès des élèves de 6^e primaire et de leurs parents. Dans les pays occidentaux, mais également dans les pays africains, il existe souvent une corrélation négative entre le niveau socio-culturel et économique de la famille et le nombre d'enfants. En effet, les familles les moins favorisées ont généralement plus d'enfants. Si l'on estime la répartition des familles selon leur niveau socio-économique sur base d'un échantillon représentatif d'élèves, on risque de surestimer le pourcentage de familles défavorisées puisque la probabilité qu'un de leur enfant soit sélectionné est plus élevée.

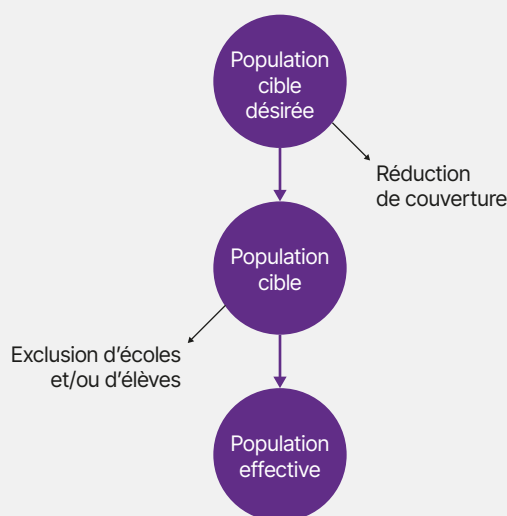
1.3. Exclusions

À côté de la population cible qui devrait inclure tous les élèves éligibles, on définit la population effective. Théoriquement, ces deux populations devraient être identiques mais compte tenu des possibles exclusions, la population effective ne recouvre pas entièrement la population cible : elle en constitue un sous-ensemble. Ces exclusions peuvent s'opérer à trois moments distincts :

1. au moment de l'opérationnalisation de la population effective : dans ce cas de figure, on parle de réduction du recouvrement de la population cible ;
2. au niveau des écoles ;
3. au niveau des élèves au sein des écoles.

La *Figure 1* illustre ces niveaux d'exclusion.

FIGURE 1 : Population cible, population effective et niveaux d'exclusion



Source : Auteurs

Réduction de « couverture » ou exclusion de la population cible

Les contraintes liées aux opérations de collecte de données peuvent conduire à exclure, dans certains pays, des zones géographiques difficilement accessibles, voire dangereuses pour les administrateurs de test en raison de conflits armés ou de terrorisme.

Des contraintes budgétaires ou politiques peuvent encore forcer la décision de ne pas évaluer des individus instruits dans une langue minoritaire dans un pays. Quelles que soient l'importance et la raison de l'exclusion, il importe de les documenter largement dans les rapports ou annexes techniques.

Enfin, les bases de sondage utilisées dans le processus d'échantillonnage peuvent être incomplètes, excluant ainsi une partie de la population sans pour autant que l'on puisse décrire ces exclusions.

Bien que fortement découragée, la réduction de la population cible n'est pas limitée dans les études de l'IEA, alors qu'elle l'est généralement dans les études PISA. Ainsi, la Lettonie dans les études PIRLS et TIMSS a pour habitude d'exclure toutes les écoles qui instruisent les élèves en Russe. Ainsi, en 2007, la sous-population instruite en russe exclue de l'enquête représentait 28%. Dans PISA, cette sous-population fait partie de la population effective.

Le PASEC a quant à lui décidé d'exclure de sa population cible les écoles qui ne relèvent pas de l'autorité nationale ou qui sont en zone d'insécurité permanente. Ainsi, les écoles internationales généralement réservées aux coopérants ne feront pas partie de la population cible et ne seront donc pas reprises dans la *Base de sondage*. Seront également automatiquement exclues de la population cible les écoles dont la langue d'enseignement n'est pas une langue nationale. Le PASEC peut être amené, sur demande expresse d'un pays, à intégrer les écoles sous expérimentation principalement sur les innovations pédagogiques notamment le bilinguisme. À cet effet, des écoles enseignant en arabe (au Tchad principalement) en wolof, bamanankan, hausa/zarma... respectivement au Sénégal, au Mali et au Niger ont été prises en compte dans le plan d'échantillonnage de ces pays. Ainsi, la population cible désirée pour l'enquête est constituée de l'ensemble des élèves de 6^e du primaire et 9^e/10^e année (fin collège) mais après réduction de couverture.

Ces exclusions feront l'objet d'intenses discussions entre les responsables nationaux et les responsables internationaux. En effet, il importe de ne pas exclure trop rapidement des sous-populations, que ce soit pour des raisons théoriques, politiques ou pratiques car, selon leur importance, de telles exclusions altèrent *in fine* la comparabilité des données. Ainsi, comment peut-on défendre la comparabilité des résultats si dans un pays, l'exclusion de l'enseignement privé concerne seulement 5 % des élèves alors que dans un autre pays, cette exclusion concerne 20 %.

Exclusions de la population effective

Certains élèves peuvent présenter des dysfonctionnements intellectuels ou physiques permanents qui les empêchent de répondre à un test ou un questionnaire, nécessitant ainsi leur exclusion. Il importe à ce niveau de clairement différencier :

1. dysfonctionnements intellectuels et difficultés scolaires. Un élève faible mais qui ne présente aucun dysfonctionnement intellectuel profond ne peut être exclu sur cette base ;
2. dysfonctionnement moteur permanent ou temporaire. Pour qu'un élève soit exclu sur cette base, il faut que le handicap soit permanent et lourd¹. Ainsi, un élève avec un bras cassé ne sera pas exclu, mais il sera considéré comme élève absent et il ne sera pas remplacé.

Enfin, les élèves étrangers qui fréquentent l'enseignement depuis moins d'un an et pour lesquels la langue maternelle diffère de la langue ou des langues d'instruction seront également exclus.

Dans certains pays, les élèves qui présentent des dysfonctionnements intellectuels ou moteurs lourds fréquentent des établissements qui se sont spécialisés dans l'accueil de tels élèves. Dans ce cas de figure, l'école entière sera exclue de population effective.

Lorsqu'un tel élève fréquente un établissement ordinaire, il sera intégré à la liste d'élèves qui sert de base à l'échantillonnage au sein des écoles, mais s'il figure parmi les 28 élèves échantillonnés (6^e année du primaire), il sera exclu et il ne sera pas remplacé par un autre élève. Il est en effet préférable de ne pas exclure au sein des écoles ordinaires les élèves qui présentent des dysfonctionnements permanents lourds avant le tirage des élèves afin de limiter autant que faire se peut ce type d'exclusion.

Il peut arriver enfin, à titre exceptionnel et pour autant que cela ne représente pas plus de 1% de la population élève, qu'une école soit éliminée sur base individuelle pour une des deux raisons suivantes :

1. l'école est très petite, c'est-à-dire qu'elle comporte moins de 8 élèves régulièrement inscrits dans le niveau scolaire considéré ;
2. l'école est géographiquement éloignée (même si la zone elle-même n'a pas été exclue).

Les exclusions de la population effective pour les raisons susmentionnées doivent être impérativement limitées à 5 % pour ne pas mettre en péril la comparabilité internationale.

Quelles qu'en soient les raisons, quel que soit le niveau auquel elles s'opèrent, la réduction de la population cible et/ou les exclusions doivent être quantifiées et largement décrites dans les rapports ou annexes techniques.

Il est possible d'estimer, sur base de quelques présupposés, le biais engendré par l'exclusion et par la non-réponse. Ainsi, si la population se distribue normalement, et pour autant que les élèves exclus soient les élèves les moins performants, la moyenne de la population sera égale à :

$$E(x) = \int_x^{+\infty} [xf(x)]dx = \frac{\phi(x)}{(1 - \Phi(x))}$$

avec :

$$\phi(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right], \text{ la fonction de densité de la distribution normale et } \Phi(x) \text{ la fonction de répartition de la}$$

distribution normale. Supposons que la performance des élèves dans une discipline donnée se distribue normalement avec une moyenne de 500 et un écart-type de 100. La figure suivante illustre cet exemple. La zone rouge représente 5 %. Le dénominateur peut donc être facilement calculé. En effet, $\Phi(x) = 5$, donc le dénominateur sera égal à 0.95.

¹ Les élèves présentant un handicap léger ou des problèmes d'apprentissage ne sont pas exclus de l'enquête.

Pour estimer la valeur du numérateur, il faut calculer la valeur de x dans la *Figure 2*. Cette valeur, égale à $-1,645$, peut être calculée au départ d'une table de répartition d'une distribution normale réduite ou en utilisant dans Excel la fonction « =LOI.NORMALE.STANDARD.INVERSE.N(0,05) » avec le 0.05 entre parenthèse représentant les 5 % d'exclusions.

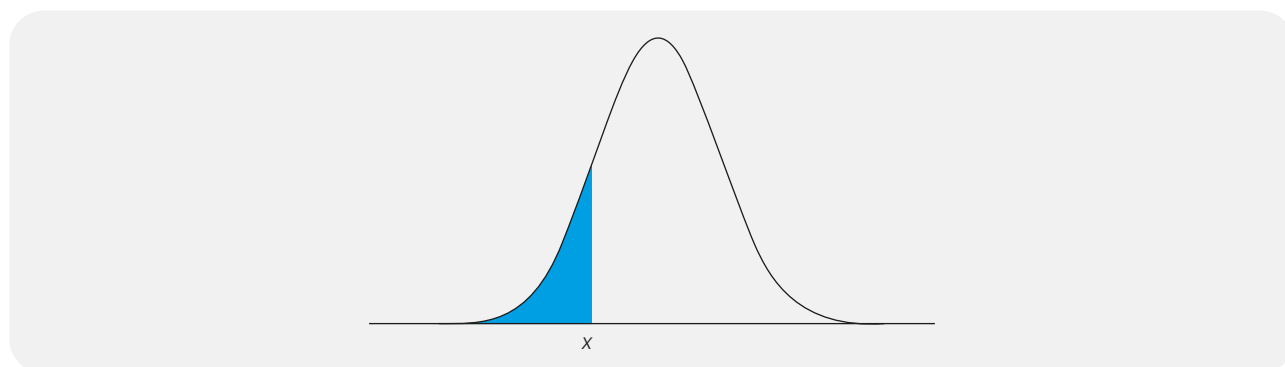
Il faut ensuite calculer :

$$\phi(-1.645) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}(-1,645)^2\right] = 0.1085$$

Cette valeur correspond à la moyenne tronquée d'une variable normale réduite si les 5 % d'élèves les plus faibles sont exclus. Pour obtenir la valeur sur l'échelle 500, 100, il suffit de multiplier la moyenne tronquée par 100 et d'ajouter 500 comme suit :

$$[(0,1085)(100)] + 500 = 510,85$$

FIGURE 2 : Exclusion des élèves les plus faibles



Le *Tableau 1* présente les moyennes tronquées en fonction du pourcentage d'exclusion, sous l'hypothèse d'une distribution normale de la performance de l'ensemble de la population élèves et sous l'hypothèse de l'exclusion des élèves les plus faibles.

TABEAU 1 : Moyennes tronquées selon le pourcentage d'exclusion

Pourcentage d'exclusion	Moyenne tronquée
0,050	510,9
0,100	519,5
0,150	527,4
0,200	535,0
0,250	542,4
0,300	549,7
0,350	557,0
0,400	564,4
0,450	572,0
0,500	579,8

Ainsi, l'exclusion des 5 % d'élèves représente une surestimation d'un dixième d'écart type, de 10 %, de deux dixième d'écart type. Bien entendu, ces moyennes tronquées donnent les valeurs maximales des biais puisque ce sont les élèves les plus faibles qui sont exclus. Mais le *Tableau 1* montre toute l'importance de minimiser les exclusions afin de maintenir une certaine comparabilité internationale.

1.4. Plans d'échantillonnage et pondération

Toute mesure est entachée d'une inéluctable erreur de mesure, variable et dans une certaine mesure contrôlable, selon la qualité et la précision de l'instrument de mesure net selon la plus ou moins grande standardisation des conditions de passation des épreuves.

Réaliser une collecte de données par échantillonnage introduit un facteur d'erreur supplémentaire, dite erreur d'échantillonnage (Neyman, 1934 ; Kish & Hess, 1959 ; Durbin, 1967). Cette erreur représente l'incertitude dans les résultats statistiques de l'enquête imputables aux différences qui existent entre les échantillons possibles que l'on peut extraire de la population de référence. En effet, même au départ d'une population de taille réduite N , il existe un grand nombre d'échantillons de taille n et selon l'échantillon retenu, la moyenne ou l'écart-type ou tout autre indicateur statistique variera plus ou moins. Ainsi, à titre d'exemple, au départ d'une population de 36 individus, il existe un peu moins de 10 milliards d'échantillons de taille 18. Le plan d'échantillonnage et la taille de l'échantillon devront donc être adéquatement déterminés afin de minimiser et contrôler au maximum ce facteur d'erreur. Une étude sera d'autant plus précise que l'erreur d'échantillonnage est faible.

Il existe différentes méthodes pour prélever un échantillon. Il est essentiel que tout chercheur comprenne bien le plan d'échantillonnage utilisé pour collecter les données sur lesquelles il ou elle travaille. Sans compréhension des propriétés du plan d'échantillonnage, les analyses peuvent être biaisées et entraîner des inférences incorrectes.

On distingue d'abord deux types d'échantillonnage, basé soit sur une sélection non probabiliste, soit sur une sélection probabiliste. L'échantillonnage non-probabiliste n'est pas envisagé ici car il s'agit d'un échantillonnage avec sélection arbitraire des éléments, ne fournissant aucune garantie quant à la représentativité des résultats (Neyman, 1934). Ce type d'échantillonnage ne constitue pas une approche rigoureuse et ne permet pas de faire des inférences correctes au sujet d'une population. Un échantillon de convenance peut toutefois être utilisé dans le cadre de la mise à l'essai des instruments et des procédures. On veillera néanmoins à garantir une certaine diversité dans cet échantillon de convenance car valider les outils et procédures sur un public d'élèves essentiellement favorisé ou essentiellement défavorisé pourrait conduire à des estimations biaisées des corrélations bisérialles de point à titre d'exemple.

L'échantillonnage probabiliste est la première base fondamentale de toute enquête de qualité. Il repose sur le principe que tout individu de la population cible a une probabilité connue et non nulle d'appartenir à l'échantillon. Si, par ailleurs, tous les individus ont la même probabilité d'être sélectionnés, l'échantillon est dit équiprobabiliste.

Suivant les caractéristiques de la population et les objectifs de l'étude, le plan d'échantillonnage sera néanmoins différent. Cependant dans tous les cas, il faut s'assurer d'éviter tout biais dans la sélection :

- l'échantillonnage doit reposer sur une sélection aléatoire. Si la sélection est consciemment ou inconsciemment influencée par des choix humains, la sélection des individus sera potentiellement biaisée (Neyman, 1934) ;
- la *Base de sondage* (en anglais, *Sampling frame*) qui sert de référent pour la sélection des individus doit couvrir au mieux l'ensemble de la population de l'étude.

Un biais peut également survenir si certaines parties de la population ne sont pas joignables et donc impossibles à évaluer, ou si certains groupes refusent de participer (Ardilly, 2006). Les exclusions ou la non-participation sont sources de biais potentiels. L'importance d'un biais pour non-réponse ou exclusion est fonction du caractère non aléatoire de celles-ci, le biais étant d'autant plus important que les non-réponses ou les exclusions sont liées à l'objet d'étude. Le biais est aussi fonction du taux d'exclusions et/ou de non-réponse. Si ce taux est élevé, plus l'ampleur du biais risque d'être élevée.

Dans le cadre des enquêtes internationales en sciences de l'éducation, les écoles peuvent refuser de participer à une étude et au sein des écoles participantes, certains élèves peuvent refuser de participer ou simplement être absents le jour de l'évaluation. Ainsi, par exemple, si les élèves faibles sont davantage absents que les élèves forts le jour d'une évaluation cognitive, la non-réponse de ces élèves introduira un biais par surestimation des résultats. Par contre, pour autant que la maladie soit considérée comme un phénomène aléatoire qui touche indistinctement les élèves selon leur performance scolaire ou selon leur origine sociale, la non-réponse pour cause de maladie devrait, dans ces conditions, n'engendrer aucun biais. Toutefois, dans le contexte africain, en raison des conditions de vie des familles défavorisées et de leur accessibilité aux soins médicaux, il est probable que la maladie touche différemment les élèves selon leur origine, ce qui engendrerait un biais dans les résultats.

Pour limiter la taille du biais dû à la non-réponse, les études internationales en sciences de l'éducation fixent un taux minimal de participation des écoles et des élèves. Pour qu'un échantillon PISA soit validé sans restriction, il faut que 85 % des écoles échantillonnées et 80 % des élèves prennent part à l'évaluation. Selon les études, les règles d'adjudication des échantillons varient sensiblement mais les principes majeurs sous-jacents restent identiques. En outre, toute étude de qualité doit s'efforcer de récolter un maximum d'informations sur les caractéristiques des non-répondants, en vue d'estimer le biais potentiel.

Pour des raisons liées aux modalités et/ou aux contraintes de la démarche d'échantillonnage, les éléments de la population peuvent ne pas avoir les mêmes chances de sélection et donc ne pas satisfaire au principe fondamental de sélection équiprobable. Si les paramètres de population sont estimés sans tenir compte de ces variabilités de probabilité de sélection, les résultats risquent également d'être biaisés (Neyman, 1934 ; Kish & Hess, 1959). Pour compenser ces probabilités variables, il importe de pondérer les données. La pondération consiste à attribuer un « poids » à chaque élément de l'échantillon qui est fonction de sa probabilité de sélection dans l'échantillon. Ce poids exprime en quelque sorte le nombre d'éléments de la population représentés par chaque élément de l'échantillon. Il s'agit dès lors de reconnaître que certains éléments sont sous-représentés dans l'échantillon et doivent donc contribuer davantage que les autres dans l'estimation des paramètres de population. Une unité d'échantillonnage dont la probabilité de sélection est faible sera considérée comme plus importante qu'une unité d'échantillonnage avec une forte probabilité de sélection. Ainsi, les poids sont inversement proportionnels à la probabilité de sélection.

Dans les sections qui suivent, différents plans d'échantillonnage vont être présentés. Pour bien comprendre les choix qui seront faits dans le cadre d'une étude internationale comme celle du PASEC, les deux principes majeurs de l'échantillonnage expliqués ci-avant doivent être gardés en mémoire : (i) le plan d'échantillonnage doit éviter tout biais dans la procédure de sélection et (ii) le plan d'échantillonnage doit permettre d'atteindre un maximum de précision en accord avec les moyens financiers alloués pour l'étude.

L'échantillonnage aléatoire et simple

En statistique inférentielle, l'échantillon aléatoire simple constitue la référence (en anglais, *Simple Random Sample* ou SRS) (Kish, 1965 ; Cochran, 1977). D'ailleurs, les logiciels de statistiques tels que SAS, SPSS, STATA ou STATISTICA, lorsqu'ils recourent à un test d'inférence ou lorsqu'ils calculent une erreur-type, ils présupposent que les individus auprès desquels les données ont été récoltées ont été sélectionnés selon une procédure aléatoire simple (IBM Corporation, 2021 ; Dove et al., 2020 ; StataCorp. 2019 ; Binder, 1983).

Trois conditions doivent être rencontrées pour obtenir un échantillon aléatoire simple :

1. chaque individu de la population a une probabilité connue non nulle d'être sélectionné ;
2. chaque individu a la même probabilité d'être sélectionné. L'échantillon est donc équiprobabiliste ;
3. les individus sont sélectionnés de manière indépendante.

Il y a plusieurs manières de tirer un tel échantillon aléatoire simple :

- les N^2 membres de la population sont numérotés et n d'entre eux sont sélectionnés par tirage aléatoire de n numéros sans remplacement ;
- N disques numérotés sont placés dans une caisse, bien mélangés, et n d'entre eux sont sélectionnés aléatoirement ;
- les N membres de la population sont ordonnés de façon aléatoire, et chaque N/n^e membre est sélectionné ;
- un nombre aléatoire est assigné à chaque N membre de la population. Les nombres aléatoires sont classés par ordre croissant. Les n premiers membres font partie de l'échantillon.

2 N représente la taille de la population et n la taille de l'échantillon.

Si n éléments sont sélectionnés à partir d'une population de N éléments selon une procédure aléatoire simple, alors la probabilité pour chaque élément j de faire partie de l'échantillon est :

$$p_j = \frac{n}{N}$$

Par exemple, si 40 élèves sont sélectionnés aléatoirement à partir d'une population de 400 élèves, la probabilité pour chaque élève j de faire partie de l'échantillon est égale à :

$$p_j = \frac{n}{N} = \frac{40}{400} = 0.10$$

En d'autres termes, chaque élève a une chance sur dix d'être sélectionné.

Le poids attribué à un individu de l'échantillon est égal à l'inverse de sa probabilité de sélection. Dans un échantillon aléatoire simple, le poids sera donc égal à :

$$w_j = \frac{1}{p_j} = \frac{N}{n}$$

Le poids de chacun des 40 élèves sélectionnés à partir d'une population de 400 élèves sera donc égal à :

$$w_j = \frac{1}{p_j} = \frac{N}{n} = \frac{400}{40} = 10$$

Cela veut dire que chaque élève de l'échantillon représente lui-même ainsi que 9 autres élèves. Comme chaque unité a la même probabilité de sélection, le poids affecté à chaque unité de l'échantillon sera identique également.

La somme des poids des éléments sélectionnés sera égale à la taille de la population, c'est-à-dire N .

$$\sum_{j=1}^n w_j = \sum_{j=1}^n \frac{N}{n} = N$$

Dans l'exemple,

$$\sum_{j=1}^{40} 10 = 400$$

Cependant, comme chaque unité de l'échantillon a le même poids, l'estimation de tout paramètre de population ne sera pas affectée par la pondération. Par exemple, considérons la moyenne d'une caractéristique appelée X . La moyenne pondérée est égale à la somme des produits du poids par la variable X et divisé par la somme des poids.

$$\hat{\mu}_{(X)} = \frac{\sum_{j=1}^n w_j x_j}{\sum_{j=1}^n w_j}$$

Comme w_j est constant, la moyenne pondérée et la moyenne non pondérée seront égales.

$$\hat{\mu}_{(X)} = \frac{\sum_{j=1}^n w_j x_j}{\sum_{j=1}^n w_j} = \frac{w_j \sum_{j=1}^n x_j}{w_j \sum_{j=1}^n 1} = \frac{\sum_{j=1}^n x_j}{n}$$

Cependant, même avec un échantillon équiprobable, les logiciels statistiques peuvent générer des résultats différents pour les données pondérées et non pondérées, notamment lors du calcul de statistiques impliquant un calcul de la variance (Binder, 1983). Le logiciel SAS, par exemple, propose quatre options pour diviser la somme des carrés et par défaut, il divise la somme des carrés pondérés par $(n-1)$ tandis que SPSS utilise comme diviseur la somme des poids moins - 1.

Soit un échantillon équiprobable de 10 individus. Le *Tableau 2* présente la taille de ces 10 individus de même pondération.

TABLEAU 2 : Taille et pondération de 10 individus

Individu	Pondération	Taille
1	10	160
2	10	162
3	10	164
4	10	166
5	10	168
6	10	170
7	10	172
8	10	174
9	10	176
10	10	178

Source : Auteurs

TABLEAU 3 : Estimation pondérée et non pondérée de l'écart-type

	Estimation de l'écart-type
SAS© estimation non pondérée	6.0553
SAS© estimation pondérée, option par défaut	19.14854
SAS© estimation pondérée, option = N, <i>nbre d'observations</i>	18.1659
SAS© estimation pondérée, option = DF, <i>nbre d'observations - 1</i>	19.14854
SAS© estimation pondérée, option = WGT, <i>somme des poids des N obs</i>	5.74456
SAS© estimation pondérée, option = WDF, <i>somme des poids des N obs - 1</i>	5.7735
SPSS© estimation non pondérée	6.0553
SPSS© estimation pondérée	5.7735

Source : Auteurs

Le Tableau 3 indique clairement comment une estimation de population peut être affectée par le processus de pondération utilisé par un logiciel statistique, même avec un échantillon équiprobable. Il est donc fortement recommandé à tout chercheur en charge d'analyses de lire attentivement la documentation relative à la pondération, fournie avec le logiciel statistique utilisé. Pour éviter ce type d'erreurs, il est conseillé de recourir aux poids normalisés dont la somme est égale au nombre d'observations valides dans la base de données. La section relative à la pondération des données abordera la normalisation des poids.

Les plans d'échantillonnage pour les études nationales et internationales en sciences de l'éducation

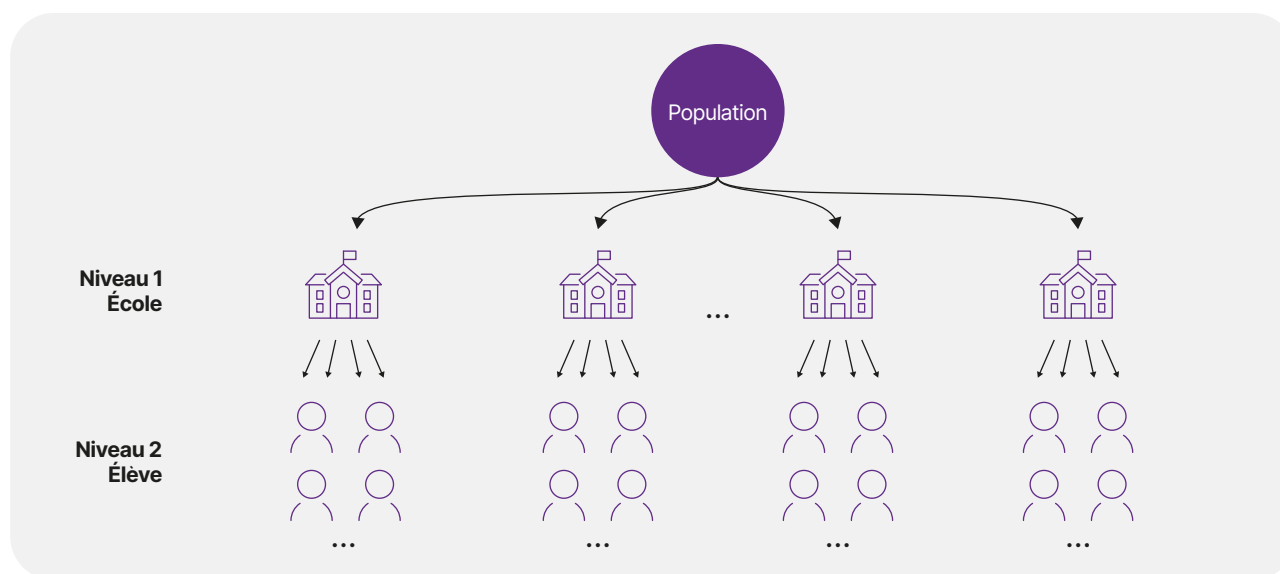
L'échantillonnage aléatoire et simple est rarement utilisé lors des études internationales en sciences de l'éducation essentiellement pour les raisons suivantes :

- ce type d'échantillonnage est très coûteux. En effet, les élèves de l'échantillon seraient très probablement répartis dans un très grand nombre d'écoles, excepté si le pays ne compte qu'un nombre limité d'écoles. Cela nécessiterait la formation et le paiement de nombreux administrateurs de test, des coûts de déplacements importants, etc. ;
- suivant la même logique, un SRS n'est pas pratique et nécessiterait de contacter beaucoup trop d'écoles ;
- d'un point de vue statistique, il ne permet pas de lier les variables élèves aux variables école, classe ou encore enseignant.

Les études en sciences de l'éducation procèdent généralement à un échantillonnage en plusieurs niveaux ou degrés (Hansen & Hurwitz, 1943 ; Horvitz & Thompson, 1952 ; Kish & Hess, 1959). Dans la plupart des cas, le premier niveau consiste en la sélection d'un échantillon d'écoles à partir de la liste complète des écoles instruisant les élèves de la population cible. Ensuite suivant les objectifs de l'étude, au deuxième degré, un échantillon aléatoire et simple d'élèves ou de classes est prélevé dans chacune des écoles sélectionnées. Un troisième degré d'échantillonnage peut être défini lorsqu'un échantillon aléatoire d'élèves de la classe est prélevé.

La Figure 3 ci-dessous illustre un échantillon en grappes à deux niveaux.

FIGURE 3 : Échantillon en grappes à deux niveaux



Dans certains pays tels que les États-Unis ou la Russie, il est possible de choisir comme premier niveau non pas les écoles mais des régions et ensuite, comme deuxième niveau, des écoles. Cependant, un tel plan d'échantillonnage nécessite qu'un certain nombre de régions soient sélectionnées (généralement au moins une quarantaine).

Échantillonnage en grappes à deux niveaux

L'étude PISA procède à un échantillonnage à deux degrés (niveaux) : un échantillon d'écoles suivi d'un échantillon aléatoire d'élèves de 15 ans dans les écoles sélectionnées (Adams & Wu, 2003). L'étude TIMSS échantillonne des écoles et ensuite des classes de grade 4 (ou grade 8 car cette étude cible deux populations) (Martin et al., 2003). Tous les élèves des classes sélectionnées font partie de l'échantillon final d'élèves (Hansen & Hurwitz, 1943 ; Horvitz & Thompson, 1952 ; Kish, 1965 ; Cochran, 1977 ; Särndal et al., 1992).

Les évaluations du PASEC établissent un plan d'échantillonnage à trois degrés : une sélection d'écoles au sein desquelles une classe de 6^e année au primaire (respectivement 9^e /10^e au collège) et pour certaines d'entre-elles une classe de 2^e année au primaire sont aléatoirement sélectionnées. Enfin 28 et 30 élèves inscrits dans la classe de 6^e année au primaire et 9^e/10^e du collège respectivement retenue sont sélectionnés selon la même procédure aléatoire simple. Dans la classe de 2^e année au primaire, ce sont 16 élèves qui sont sélectionnés.

Supposons que la population de 400 élèves évoquée dans le paragraphe précédent est répartie dans 10 écoles, chaque école comptabilisant 40 élèves, ni plus, ni moins. Quatre écoles sont sélectionnées aléatoirement et dans chacune des quatre écoles sélectionnées, 10 élèves sont sélectionnés selon la même procédure. Chaque école, notée i , a une probabilité de sélection de :

$$p_i = \frac{n_{sc}}{N_{sc}} = \frac{4}{10} = 0.4$$

avec N_{sc} étant le nombre d'écoles de population et n_{sc} le nombre d'écoles échantillonnées.

Dans les quatre écoles sélectionnées, chaque élève, noté j , a une probabilité de sélection égale à :

$$p_{j/i} = \frac{n_i}{N_i} = \frac{10}{40} = 0.25$$

avec N_i étant le nombre d'élèves dans l'école i et n_i le nombre d'élèves sélectionnés dans l'école i . Cela veut dire que dans chaque école sélectionnée, chaque élève a une chance sur 4 d'être sélectionné.

La probabilité finale de sélection pour l'élève j dans l'école i est égale au produit de la probabilité de sélection de l'école i par la probabilité de sélection de l'élève j dans l'école i (Hansen & Horwitz, 1943), c'est-à-dire :

$$p_{ji} = p_i * p_{j/i} = \frac{n_{sc} n_i}{N_{sc} N_i}$$

Dans l'exemple, la probabilité finale de sélection de l'élève est égale à :

$$p_{ji} = p_i * p_{j/i} = \frac{n_{sc} n_i}{N_{sc} N_i} = \frac{4 * 10}{10 * 40} = 0.4 * 0.25 = 0.10$$

Le poids école, noté w_i , le poids élève dans l'école, noté $w_{j/i}$, et le poids final de l'élève, noté w_{ji} , sont respectivement égaux à :

$$w_i = \frac{1}{p_i} = \frac{1}{0.4} = 2.5$$

$$w_{j/i} = \frac{1}{p_{j/i}} = \frac{1}{0.25} = 4$$

$$w_{ji} = \frac{1}{p_{ji}} = \frac{1}{0.1} = 10$$

Le *Tableau 4* présente la probabilité de sélection de l'école, de l'élève dans l'école et la probabilité finale de l'élève ainsi que les poids respectifs à ces différents niveaux si dans notre exemple, les écoles 2, 5, 7 et 10 ont été reprises dans l'échantillon.

Le *Tableau 4* montre que la somme des poids école correspond au nombre d'écoles dans la population, c'est-à-dire 10, et la somme des poids élèves finaux correspond au nombre d'élèves dans la population, c'est-à-dire 400. La pondération de l'échantillon permet donc de reconstruire la population. Ce tableau montre également que chaque élève a la même probabilité finale d'être sélectionné. Ces élèves auront donc tous le même poids. L'échantillon est dit équi-probabiliste et dès lors auto-pondéré.

TABLEAU 4 : Probabilités de sélection et poids correspondants pour un échantillonnage à deux niveaux, avec échantillonnage aléatoire et simple au premier niveau école, les écoles étant de tailles égales

Numéro de l'école	Taille de l'école N_i	Probabilité école p_i	Poids école w_i	Probabilité élève dans l'école p_{ji}	Poids élève dans l'école w_{ji}	Probabilité finale de l'élève p_{ji}	Poids final de l'élève w_{ji}	Somme des poids finaux $n_i w_{ji}$
1	40							
2	40	0.4	2.5	0.25	4	0.1	10	100
3	40							
4	40							
5	40	0.4	2.5	0.25	4	0.1	10	100
6	40							
7	40	0.4	2.5	0.25	4	0.1	10	100
8	40							
9	40							
10	40	0.4	2.5	0.25	4	0.1	10	100
Total	400		10.0					400

Source : Auteurs

En pratique évidemment, la taille des écoles varie. Ainsi, il y a en général plus d'élèves par école dans les zones urbaines que dans les zones rurales. Si les écoles sont sélectionnées par SRS, la probabilité de sélection des écoles est identique mais au sein des écoles sélectionnées, la probabilité de sélection des élèves variera en fonction de la taille de l'école. Dans une petite école, la probabilité pour un élève d'être sélectionné sera grande tandis que dans une grande école, cette probabilité sera faible. Il s'ensuit que l'échantillon d'élèves ne peut plus être considéré comme un échantillon équi-probabiliste. Le Tableau 5 montre un exemple de résultats obtenus par SRS lorsque la taille des écoles varie.

TABLEAU 5 : Probabilités de sélection et poids correspondants pour un échantillonnage à deux niveaux, avec échantillonnage aléatoire et simple au premier niveau école, la taille des écoles variant

Numéro de l'école	Taille de l'école N_i	Probabilité école p_i	Poids école w_i	Probabilité élève dans l'école p_{ji}	Poids élève dans l'école w_{ji}	Probabilité finale de l'élève p_{ji}	Poids final de l'élève w_{ji}	Somme des poids finaux $n_i w_{ji}$
1	10							
2	15	0.4	2.5	0.67	1.5	0.27	3.75	37.5
3	20							
4	25							
5	30	0.4	2.5	0.33	3	0.13	7.5	75
6	35							
7	40	0.4	2.5	0.25	4	0.10	10	100
8	45							
9	80							
10	100	0.4	2.5	0.10	10	0.04	25	250
Total	400		10					462.5

Source : Auteurs

Selon une sélection aléatoire et simple d'écoles de tailles différentes, toutes les écoles ont la même probabilité de sélection et comme précédemment, la somme des poids école sera égale au nombre d'écoles dans la population. Par contre, la somme des poids finaux élève ne sera pas nécessairement égale au nombre d'élèves dans la population. De plus, le poids final de l'élève variera selon la taille de l'école.

Cette variabilité réduira la validité de toutes les estimations des paramètres de population.

Les Tableaux 6 et 7 présentent respectivement les probabilités et les poids lorsque ce sont les quatre plus petites écoles qui sont sélectionnées et lorsque ce sont les quatre plus grandes.

TABLEAU 6 : Probabilités de sélection et poids correspondants pour un échantillonnage à deux niveaux, avec SRS au niveau école – Sélection des quatre plus petites écoles

Numéro de l'école	Taille de l'école N_i	Probabilité école p_i	Poids école w_i	Probabilité élève dans l'école $p_{j i}$	Poids élève dans l'école $w_{j i}$	Probabilité finale de l'élève p_{ji}	Poids final de l'élève w_{ji}	Somme des poids finaux $n_i w_{ji}$
1	10	0.4	2.5	1.00	1	0.40	2.5	25
2	15	0.4	2.5	0.67	1.5	0.27	3.75	37.5
3	20	0.4	2.5	0.50	2	0.20	5	50
4	25	0.4	2.5	0.40	2.5	0.16	6.25	62.5
Total			10					175

Source : Auteurs

TABLEAU 7 : Probabilités de sélection et poids correspondants pour un échantillonnage à deux niveaux, avec SRS au niveau école – Sélection des quatre plus grandes écoles

Numéro de l'école	Taille de l'école N_i	Probabilité école p_i	Poids école w_i	Probabilité élève dans l'école $p_{j i}$	Poids élève dans l'école $w_{j i}$	Probabilité finale de l'élève p_{ji}	Poids final de l'élève w_{ji}	Somme des poids finaux $n_i w_{ji}$
7	40	0.4	2.5	0.25	4	0.10	10	100
8	45	0.4	2.5	0.22	4.5	0.09	11.25	112.5
9	80	0.4	2.5	0.13	8	0.05	20	200
10	100	0.4	2.5	0.10	10	0.04	25	250
Total			10					662.5

Source : Auteurs

Si le hasard de la sélection entraîne que seules des petites écoles sont sélectionnées, la population élèves sera largement sous-estimée. À l'inverse si seules les grandes écoles font partie de l'échantillon, la taille de la population élève sera largement surestimée. Le SRS d'écoles permet de reconstruire exactement la population école mais pas nécessairement la population élève.

Or, les études internationales en sciences de l'éducation telles que les études du PASEC ou PISA s'intéressent davantage à l'échantillon d'élèves qu'à l'échantillon d'écoles. L'échantillon d'écoles pourrait en quelque sorte être considéré comme un niveau opérationnel en vue de tirer l'échantillon d'élèves. On peut raisonnablement conclure qu'un plan d'échantillonnage passant par un SRS d'écoles est inapproprié puisqu'il peut largement sous-estimer ou surestimer la taille de la population d'élèves. Il en résulte également une importante variation des poids finaux des élèves selon la taille de l'école. Cette variation de poids est susceptible d'augmenter de façon conséquente la variance d'échantillonnage.

Pour éviter ces inconvénients, les études internationales en sciences de l'éducation privilégient la sélection d'écoles avec une probabilité proportionnelle à leur taille (en anglais, *probability proportional to their size* ou PPS) (Adams & Wu, 2003 ; Martin et al., 2003 ; Hansen & Hurwitz, 1943). Cela signifie que les écoles de grande taille ont une plus grande probabilité de sélection que les plus petites écoles. Par contre, les élèves des grandes écoles auront une plus petite probabilité intra-école d'être sélectionnés que les élèves des petites écoles.

Dans le cas d'un échantillon PPS, la probabilité pour une école d'être sélectionnée est égale au produit de la taille de l'école N_i par le nombre d'écoles à sélectionner n_{sc} , le tout divisé par le nombre total d'élèves dans la population N :

$$p_i = \frac{N_i n_{sc}}{N}$$

Les formules pour calculer les probabilités et les poids intra-école restent inchangés. De même, la probabilité finale et le poids final sont toujours égaux au produit des probabilités (ou poids) école par les probabilités (ou poids) intra-école. Dans notre exemple, la probabilité de sélection de l'école 9 est égale à :

$$p_9 = \frac{N_9 * n_{sc}}{N} = \frac{80 * 4}{400} = \frac{4}{5} = 0.8$$

La probabilité intra-école pour l'élève j dans l'école 9 sera égale à :

$$p_{j/9} = \frac{n_9}{N_9} = \frac{10}{80} = 0.125$$

La probabilité finale de l'élève j dans l'école 9 est égale à :

$$p_{j9} = 0.8 * 0.125 = 0.1$$

TABLEAU 8 : Probabilités de sélection et poids correspondants pour un échantillonnage à deux niveaux, avec PPS au niveau école

Numéro de l'école	Taille de l'école p_i	Probabilité école w_i	Poids école p_{ji}	Probabilité élève dans l'école w_{ji}	Poids élève dans l'école p_{ji}	Probabilité finale de l'élève w_{ji}	Poids final de l'élève $n_i w_{ji}$	Somme des poids finaux $n_i w_{ji}$
1	10							
2	15							
3	20	0.2	5.00	0.50	2	0.10	10	100
4	25							
5	30							
6	35							
7	40	0.4	2.50	0.25	4	0.10	10	100
8	45							
9	80	0.8	1.25	0.13	8	0.10	10	100
10	100	1	1.00	0.10	10	0.10	10	100
Total	400		9.75					400

Source : Auteurs

Comme le montre le *Tableau 8*, les poids école et intra-école sont différents selon les écoles, mais les poids finaux des élèves sont à présent constants. De plus, la somme des poids finaux des élèves correspond au nombre total d'élèves dans la population : l'échantillonnage et la pondération reconstruisent exactement la population élève. Cependant, la somme des poids école ne correspond pas exactement au nombre d'écoles dans la population. Dans l'exemple, la population école n'est pas reconstruite, elle est sous-estimée (9.75 et non 10). Dans une certaine mesure, ceci n'est pas un problème majeur, puisque la population d'intérêt est constituée d'abord des élèves et non des écoles.

Avec un échantillon PPS d'écoles et un nombre constant d'élèves SRS dans chaque école sélectionnée, la somme des poids finaux des élèves est toujours égale au nombre d'élèves dans la population (non-réponses ou erreurs dans la *Base de sondage* ignorées à ce stade), l'échantillon est alors auto-pondéré (Hansen & Hurwitz, 1943). Ce serait le cas même si l'échantillon était uniquement constitué des plus petites ou uniquement des plus grandes écoles. Dans ce cas, cependant la somme des poids école différerait très fortement de la population école. Dans notre exemple, si les 4 plus petites écoles étaient sélectionnées, la somme des poids école serait égale à 25.67 (au lieu de 10) et si les 4 plus grandes écoles avaient été sélectionnées, la somme des poids école serait égale à 6.97 (toujours au lieu de 10).

Pour éviter une trop grande différence entre le nombre d'écoles dans la population et la somme des poids école, les écoles sont sélectionnées selon une procédure systématique. Cette procédure consiste à trier les écoles selon leur taille et ensuite sélectionner une école toutes les x unités, x étant le pas d'échantillonnage. Le *Tableau 9* ci-dessous montre les 4 écoles échantillonnées au départ d'une population de 10 écoles de tailles différentes, selon une procédure systématique.

La procédure systématique implique les étapes suivantes :

1. Trier les écoles selon leur taille.
2. Définir le pas d'échantillonnage. Il s'agit du rapport entre le nombre d'élèves dans la population et le nombre d'écoles dans l'échantillon.

$$Int = \frac{N}{n_{sc}} = \frac{400}{4} = 100$$

3. Tirer un nombre aléatoire issu d'une distribution uniforme $[0;1]$. Imaginons pour notre exemple, le nombre aléatoire 0.232.
4. Ce nombre est ensuite multiplié par le pas d'échantillonnage, c'est-à-dire $0.232 \times 100 = 23.2$.
5. L'école qui contient l'élève numéro 24 (entier supérieur) est sélectionnée. Ensuite, on ajoute le pas d'échantillonnage et l'école qui contient l'élève numéro 124 est sélectionnée. Cette procédure systématique est appliquée jusqu'à ce que le nombre d'écoles à échantillonner soit atteint. Le *Tableau 9* illustre cette procédure à partir de notre exemple. Les quatre nombres sélectionnés sont : 23.2, 123.2, 223.2 et 323.2.

Trier les écoles selon la taille et ensuite procéder à une sélection systématique évite d'obtenir un échantillon de petites écoles uniquement ou, beaucoup plus probable, un échantillon de grandes écoles uniquement. En d'autres termes, cette procédure permet de tirer un échantillon qui reconstruit la population élève et qui conserve au maximum les propriétés de la population école également.

TABLEAU 9 : Sélection d'écoles par PPS selon une procédure systématique

Numéro de l'école	Taille de l'école	De l'élève numéro...	...à l'élève numéro	École échantillonnée
1	10	1	10	Non
2	15	11	25	Oui
3	20	26	45	Non
4	25	46	70	Non
5	30	71	100	Non
6	35	101	135	Oui
7	40	136	175	Non
8	45	176	220	Non
9	80	221	300	Oui
10	100	301	400	Oui

Source : Auteurs

Substance d'une variation des poids

Comme expliqué ci-dessus, un plan d'échantillonnage à deux niveaux avec un échantillonnage PPS des écoles au premier niveau et la sélection d'un nombre fixe d'élèves au sein de chacune des écoles selon une procédure SRS devrait garantir que ces derniers auront la même probabilité de sélection et donc le même poids final. Pourtant, dans la pratique, les poids finaux des élèves ne sont pas tout à fait constants. Différents facteurs contribuent à cette variabilité des poids finaux :

- **le sur-échantillonnage ou sous-échantillonnage de certaines strates de la population.** Généralement, la population d'écoles est divisée en différents sous-groupes, appelés strates (voir la section consacrée à la stratification). Par exemple, un pays peut distinguer les écoles urbaines des écoles rurales dans le listing des écoles. Dans la plupart des cas, le nombre d'élèves sélectionnés dans la strate rurale et dans la strate urbaine sera proportionnel à la taille de cette strate dans la population entière (Neyman, 1943). Mais lorsqu'une strate est composée d'un très petit nombre d'élèves, on peut décider de ne pas respecter l'allocation proportionnelle et sélectionner davantage d'élèves (et donc d'écoles) afin de pouvoir disposer de suffisamment d'observations pour les comparer avec les élèves des autres groupes (Kish, 1965). Ce processus de sur-échantillonnage permet de produire des estimations valides au niveau de cette petite strate. Si le pays décide de sur-échantillonner la strate rurale, la probabilité de sélection des écoles de cette strate sera augmentée et les élèves de cette strate auront par conséquent un poids final moindre ;
- **le manque de précision ou d'actualisation des informations relatives à la taille des écoles dans la Base de sondage.** L'échantillonnage des écoles est pratiqué sur la base de la liste complète des écoles de la population ou *Base de sondage*. Pour être en mesure d'effectuer une sélection avec probabilité proportionnelle à la taille de l'école, une mesure de taille doit être incluse dans la liste. Pour les enquêtes du PASEC, il s'agit du nombre d'élèves de 6^e année du primaire ou des élèves de fin collège inscrits dans l'école³. Mais dans certains pays, le nombre d'élèves par niveau scolaire dans l'école n'est pas répertorié dans les statistiques nationales ou n'est pas disponible pour toutes les écoles. Dans ce cas, la mesure de taille est estimée en divisant le nombre total d'élèves dans l'école par le nombre de niveaux scolaires organisés au sein de cette école. De plus, même si les statistiques nationales disposent de l'information nécessaire, les données peuvent ne pas être à jour et dater d'une, deux voire plusieurs années. Dès lors, les incohérences entre la mesure de taille indiquée dans la *Base de sondage* et le nombre d'élèves réel au moment de l'évaluation génèrent une certaine variabilité des poids finaux des élèves. Reprenons notre exemple et imaginons que l'école 9 dans le tableau 8 compte 100 élèves et non 80 au moment de l'évaluation. La probabilité de sélection de l'école 9 (0,8) a été calculée antérieurement, à partir de la mesure de taille reprise dans la *Base de sondage*, soit 80. Le poids des élèves dans l'école est calculé au moment de l'évaluation, il est donc égale à 100 divisé par 10, c'est-à-dire 10 et non 8 comme prévu. Ainsi, le poids élève final sera égal à 12,5 au

3 Suivant le plan d'échantillonnage définitif (voir point 1.7), il peut s'agir du nombre d'élèves de 6^e primaire et du nombre d'élèves de 2^e primaire considérés séparément ou de la somme des élèves des 2 niveaux scolaires.

lieu de 10 comme attendu. Il s'ensuit également que la somme des pondérations ne sera plus exactement égale à 400, mais à 425 dans le cas présent ;

- **l'ajustement des poids pour non-réponse des écoles et des élèves.** Certaines écoles et au sein des écoles participantes, certains élèves peuvent refuser de participer ou être absents le jour de l'évaluation. Pour compenser cette non-réponse, le poids élève est ajusté. Par exemple, si seulement 15 élèves d'une classe hors des 28 élèves sélectionnés en 6^e année du primaire dans une école participante sont présents le jour de l'évaluation, le poids des élèves participants sera multiplié par le ratio 28/15. Comme le taux de participation varie d'une école à l'autre, les poids finaux des élèves varient également d'une école à l'autre.

Échantillonnage en grappes à 3 degrés

Pour rappel, les enquêtes du PASEC procèdent à un échantillonnage à trois niveaux :

- 1^{er} degré d'échantillonnage : un échantillon d'écoles sélectionnées avec une probabilité proportionnelle à la taille de l'école et selon une procédure systématique ;
- 2^e degré : la classe du ou des niveau(x) scolaire(s) visé(s) ou un échantillon SRS d'une classe s'il existe plus d'une classe du niveau, au sein de chaque école sélectionnée ;
- 3^e degré : un échantillon SRS d'élèves au sein de la classe sélectionnée

Les considérations théoriques développées pour l'échantillonnage PPS à deux degrés et relatives aux pondérations sont directement applicables à un échantillonnage PPS à trois degrés tel que pratiqué par le PASEC. Cependant les classes de même niveau au sein d'une école ne sont pas nécessairement de tailles strictement égales. Cette (légère) variabilité introduit une variabilité des probabilités de sélection de l'élève, et donc une légère variabilité supplémentaire des poids finaux des élèves.

Par ailleurs, en ne choisissant qu'une seule classe, la variance entre classes au sein des écoles et la variance entre école seront confondues. Les effets écoles et les effets classes ne seront pas différenciables.

1.5. Taille de l'échantillon

Une des premières questions qui se pose lors de la conception d'un plan d'échantillonnage porte sur le nombre d'individus à sélectionner. Une conception erronée et fréquente consiste à penser que la taille de l'échantillon devrait être déterminée en proportion de la population cible. Cela consisterait à dire qu'un pays fortement peuplé devrait prévoir un échantillon plus grand qu'un pays faiblement peuplé. Or, il n'en est rien. Une analogie amusante peut être faite avec une personne chargée de tester le goût d'un potage. Tester le goût d'une casserole de soupe ne nécessite pas d'utiliser un ustensile différent de celui nécessaire pour tester le goût d'une assiette de soupe ; dans les deux cas, une cuillère apportera les informations utiles.

Quels sont les éléments à prendre en considération pour déterminer la taille de l'échantillon ?

La taille de l'échantillon dépend des objectifs poursuivis par l'étude et des moyens financiers disponibles. À titre d'exemple, si des responsables politiques nationaux souhaitent comparer les performances d'élèves de deux sous-groupes dont l'un ne représente qu'une faible proportion de la population cible, il sera probablement nécessaire d'augmenter la taille de l'échantillon du petit sous-groupe.

Ensuite, la taille de l'échantillon dépend du niveau de précision attendu dans l'estimation des paramètres de population. Plus on souhaite être précis, plus la taille de l'échantillon devra être grande. Dans le cadre des enquêtes internationales conduites notamment par l'IEA, la taille de l'échantillon est régulièrement choisie pour obtenir une précision équivalente à celle qu'on obtiendrait avec un échantillon aléatoire et simple de 400 élèves. Il en est de même pour le PASEC.

Enfin, la taille de l'échantillon dépend directement du plan d'échantillonnage et des caractéristiques de la population. Comme susmentionné, les études internationales en sciences de l'éducation adoptent des plans d'échantillonnage en grappes par degrés (en anglais, *multi-stage cluster sample designs*). Or, les élèves au sein d'une école ont tendance à partager certaines caractéristiques communes et donc à être plus ressemblants que parmi la population entière. De même, les élèves d'une même classe, instruits par le même enseignant, seront plus similaires entre eux. Observer deux individus d'une même école apporte donc moins d'informations que l'observation de deux élèves d'écoles différentes. Cette perte d'information devra donc être compensée par une augmentation de la taille de l'échantillon d'autant plus importante que les élèves d'une même école se ressemblent.

Pour quantifier le niveau de ressemblance des élèves au sein des écoles, on recourt généralement au pourcentage de la variance totale qui se situe entre écoles. Autrement dit, il s'agit du rapport entre la variance école et de la somme de la variance entre écoles et de la variance à l'intérieur des écoles. Notons qu'en raison de différences liées aux degrés de liberté, la variance *Élèves*, toute école confondue, n'est pas égale à la somme de la variance *Inter Écoles* et de la variance *Intra École*. Plus le nombre d'élèves sélectionnés par école sera faible, plus la différence sera importante. Inversement, plus le nombre d'élèves sélectionnés par école est grand, plus cette différence sera petite.

Ce coefficient se dénomme également « coefficient de corrélation intra-class » – le mot *class* faisant généralement référence à la première unité d'échantillonnage et non à la classe – car il correspond à la valeur attendue de la corrélation entre deux élèves sélectionnés aléatoirement au sein de chacune des écoles de l'échantillon.

Pour atteindre un niveau de précision équivalent à celui d'un échantillon aléatoire et simple de 400 élèves, et pour permettre d'estimer des paramètres de population au niveau élève mais encore pour étudier l'effet école et parfois l'effet maître, la taille de l'échantillon doit être sensiblement revue à la hausse lorsqu'un échantillonnage en grappes par degrés est implanté.

Imaginons donc un échantillon final de 1000 élèves. Quel serait l'échantillon le plus efficace, (i) un échantillon de 100 écoles et 10 élèves par école ou (ii) un échantillon de 20 écoles et 50 élèves par école ? Le bon choix sera fonction des caractéristiques de la population et plus particulièrement de l'ampleur du coefficient de corrélation intra-class.

S'il y a très peu de variabilité entre les écoles au sein de la population, de sorte que toute la variabilité de la population se concentre au sein des écoles, chaque école pourrait à la limite être assimilée à un sous-échantillon représentatif de la population. Dans ce cas, un échantillon comptant davantage d'élèves par école soit 20 écoles et 50 élèves par école, garantirait la même précision dans l'estimation des paramètres élèves qu'un échantillon de 100 écoles et de 10 élèves par écoles. Si par contre, la population présente une grande part de variance entre écoles, chaque école ne représente à elle seule qu'une plus petite fraction de la population et il faut privilégier un plus grand échantillon d'écoles. En effet, les élèves d'une même école se ressemblent et collecter les données auprès d'un grand nombre d'élèves au sein de chacune des écoles sélectionnées n'apporte pas beaucoup d'informations.

Par ailleurs, comme ces études internationales souhaitent également estimer l'impact de certaines caractéristiques *École* et *Maître* sur la performance des élèves, il est préférable de tirer un échantillon de 100 écoles avec 10 élèves par école, qu'un échantillon de 20 écoles, avec 50 élèves par école.

Les enquêtes internationales concernent des systèmes éducatifs qui peuvent être très différents, certains présentant de grandes différences entre écoles comme la Belgique, l'Allemagne ou encore l'Autriche, d'autres très peu comme la Finlande ou la Corée du Sud. Malgré ces spécificités nationales, les tailles des échantillons école et élève ne varient pas (ou peu) d'un pays à l'autre mais sont définies une fois pour toutes quelles que soient les caractéristiques du pays.

Les enquêtes de l'IEA (TIMSS et PIRLS) fixe généralement la taille de l'échantillon école à 150 et parmi celles-ci, une classe et tous ses élèves sont sélectionnés. L'enquête PISA sélectionne 5 250 élèves répartis au sein de 150 écoles, soit 35 élèves de 15 ans par école. Le recours aux variables de stratification permet, comme il sera décrit ci-après, l'impact de la variance inter écoles sur la taille de l'échantillon et sur la précision des estimations.

L'enquête du PASEC opte :

- pour la population des élèves en fin de scolarité primaire, un échantillon standard de 200 écoles au premier niveau, une classe au second niveau et dans celle-ci, 28 élèves sont sélectionnés, ce qui correspond à un échantillon attendu de 5 600 élèves. L'échantillon de deuxième année comportera 100 écoles et 16 élèves par école sélectionnée ;

- pour la population des élèves en fin de collège (9^e/10^e année), un échantillon standard de 180 écoles au premier niveau, une classe au second niveau et dans celle-ci, 30 élèves sont sélectionnés, ce qui correspond à un échantillon représentatif national attendu de 5 400 élèves.

Ces tailles d'échantillons école et élève devraient, pour la population de fin de scolarité primaire, dans la plupart des pays, permettre d'atteindre le niveau de précision préalablement fixé, tout en prenant en considération les taux de non-réponse. Les systèmes éducatifs qui présentent peu de variance entre écoles bénéficient par conséquent d'un niveau de précision plus élevé.

1.6. Stratification

La stratification d'une population est un second procédé permettant d'augmenter la précision d'un échantillon. Elle consiste à se servir de la connaissance que l'on a de la population pour scinder ou ordonner celle-ci en groupes homogènes selon une ou plusieurs variables déterminantes, dites variables de stratification (Neyman, 1934 ; Kish, 1965 ; Cochran, 1977).

La stratification peut être explicite ou implicite et est appliquée au premier niveau d'échantillonnage, c'est-à-dire généralement au niveau école dans le cadre des *surveys* en éducation. Elle a plusieurs fonctions dont la principale est la réduction de la variance d'échantillonnage.

La **stratification explicite** consiste en la création de groupes d'écoles ou strates homogènes et distinctes selon la variable de stratification explicite définie. L'échantillonnage des écoles est ensuite effectué de manière indépendante dans chacune des strates ainsi formées, et le nombre d'écoles sélectionnées par strate est généralement proportionnel à la taille de la strate. Plus la variable de stratification corrèle avec la variable d'intérêt de l'étude, c'est-à-dire la performance en lecture et en mathématiques pour les enquêtes du PASEC, plus les échantillons de chacune des strates seront homogènes vis-à-vis de la performance et plus la variance d'échantillonnage sera ainsi réduite. L'effet des variables de stratification sur la variance d'échantillonnage peut être estimé par le carré de la corrélation.

La stratification explicite est également particulièrement utile lorsqu'on désire appliquer un plan d'échantillonnage spécifique à une sous-population, comme un sur-échantillonnage par exemple pour permettre des comparaisons infranationales à titre d'exemple. Si les responsables politiques d'un pays désirent obtenir des estimations précises relatives à une sous-population de petite taille, comme par exemple les élèves fréquentant une école dont la langue d'enseignement est marginale dans le pays, la stratification selon la langue et le sur-échantillonnage de la petite strate permettront d'atteindre une taille d'échantillon suffisante pour effectuer des analyses fiables pour cette petite sous-population.

La **stratification implicite** consiste à trier les écoles de la *Base de sondage* selon la ou les variables définies. Ce tri est opéré à l'intérieur de chaque strate explicite. La procédure systématique d'échantillonnage des écoles – une école échantillonnée à chaque pas d'échantillonnage – garantit la représentativité proportionnelle des élèves dans l'échantillon. Par exemple, considérons un système éducatif constitué d'écoles privées et d'écoles publiques. Sans stratification, la sélection des écoles selon le type de financement est tout à fait aléatoire. Un échantillon particulier pourrait contenir une proportion d'élèves issus d'écoles privées nettement supérieure à ce qu'elle représente dans la population ou à l'inverse une proportion beaucoup plus faible qu'en réalité. Si la liste d'écoles est triée par type d'école, la procédure systématique d'échantillonnage conduira à sélectionner les écoles en respectant la proportion existant dans la population.

En général, la sélection de 3 à 5 variables de stratification, explicite et implicite confondues, est suffisante. Il faut en outre s'assurer de ne pas constituer un nombre trop élevé de strates, certaines strates contenant alors très peu d'écoles. Généralement, il est conseillé qu'une strate contienne au minimum 4 écoles avant échantillonnage. En effet, un échantillon doit contenir au minimum deux écoles par strate explicite pour être en mesure de calculer une variance d'échantillonnage. Dans le cadre du PASEC, il sera nécessaire que chaque strate explicite contienne au moins 4 écoles car au sein de chaque strate explicite, la moitié des écoles sera sélectionnée pour participer à l'évaluation en deuxième. Il restera donc au moins deux écoles participantes à l'évaluation en début de scolarité primaire.

Comme susmentionné, les variables de stratification choisies doivent corrélérer avec la mesure de l'étude pour qu'elles contribuent à réduire la variance d'échantillonnage. Les évaluations internationales des systèmes éducatifs retiennent généralement les variables de stratification parmi la liste non exhaustive suivante :

- zone géographique (état, province...) ;
- urbanisation (zone rurale, zone urbaine) ;
- type d'école (publique, privée, communautaire) ;
- filière scolaire (enseignement général, enseignement professionnel) ;
- niveau socioéconomique de l'élève ;
- langue d'enseignement ;
- organisation des classes (simple vacation, double vacation, classes multigrade).

La décision d'utiliser une variable de stratification de façon explicite ou implicite dépend de la ou des raison(s) qui motive(nt) son utilisation.

Le PASEC conseille fortement à tous les pays participants d'utiliser les variables de stratification suivantes :

- **stratification explicite** : la variable « **région** » ou « **province** ». Plusieurs petites régions peuvent être regroupées au sein d'une même strate au besoin. La variable « Organisation des classes » utilisée dans des études antérieures est déconseillée pour des raisons d'instabilité. En effet, l'organisation des classes dans l'école, en simple ou en double vacation par exemple, varie souvent d'une année à l'autre. La stratification élaborée au départ de la *Base de sondage*, peut ne plus correspondre à la réalité au moment de l'évaluation en raison du décalage temporel entre l'année scolaire couverte par la *Base de sondage* et l'année scolaire de l'évaluation sur le terrain ;
- **stratification implicite** : les variables « **type d'école** » (public/privé) et « **urbanisation** » (urbain/rural).

Néanmoins, nous conseillons d'engager une discussion avec chaque pays participant pour s'assurer de la pertinence de ce choix en lien avec le système éducatif du pays et les priorités politiques nationales.

1.7. Recouvrement des échantillons de 6^e année et 2^e année du primaire

Le programme du PASEC a pour objectif principal d'évaluer le niveau de performance des élèves en compréhension de l'écrit et en mathématiques au terme de la scolarité primaire, c'est-à-dire en 6^e année du primaire. Par ailleurs, pour aider les pays à identifier les difficultés rencontrées par les élèves en début d'apprentissage, la CONFEMEN a également décidé de proposer une évaluation en 2^e année primaire. Alors que l'administration des tests et questionnaires contextuels sera collective en sixième, il est nécessaire de recourir à une administration individualisée des instruments de mesure en deuxième. Pour cette raison, la taille de l'échantillon de deuxième sera réduite comparativement à la taille de l'échantillon de sixième.

Le recouvrement des échantillons peut être envisagé selon deux plans d'échantillonnage différents avec des implications substantielles et propres à chacun : (i) tirer deux échantillons indépendants avec un maximum de recouvrement ou (ii) tirer les deux échantillons d'élèves de 6^e primaire et de 2^e primaire à partir d'un seul échantillon d'écoles.

La première alternative, à savoir le tirage de deux échantillons indépendants avec un maximum de recouvrement entre les deux, présentent les avantages suivants :

- les deux échantillons sont représentatifs de leur population respective. Les comparaisons internationales sont dès lors possibles en début et en fin de scolarité primaire ;
- le niveau de précision de l'échantillon de 6^e année du primaire est optimal. En effet, selon cette procédure, il est impossible de sélectionner pour l'échantillon de 6^e année, une école qui ne comporte pas d'élèves à ce niveau, alors que ce cas de figure n'est pas improbable avec la seconde procédure décrite ci-dessous. Il en est de même pour l'échantillon de 2^e année du primaire ;
- bien que la taille de l'échantillon de 2^e année du primaire est deux fois plus petite, la précision de l'échantillon est moindre mais les estimations de population ne sont néanmoins pas biaisées ;

- l'application d'une procédure de recouvrement maximal entre les deux échantillons limite fortement l'augmentation du nombre total d'écoles à évaluer et la majorité des 100 écoles échantillonnées pour la 2^e année du primaire feront également partie de l'échantillon de 6^e année du primaire ;
- les tâches d'échantillonnage confiées aux conseillers techniques du PASEC sont plus lourdes et plus complexes puisqu'il s'agit de tirer deux échantillons partiellement indépendants ;
- au niveau national, la gestion de l'étude est légèrement plus compliquée puisque le nombre d'écoles impliquées peut être légèrement supérieur. L'importance du non-recouvrement entre les échantillons de 2^e année et de 6^e année dépend du pourcentage d'écoles qui proposent un seul des deux niveaux cibles de l'étude PASEC.

Cependant, le tirage des échantillons élève pour les deux niveaux scolaires à partir d'un seul échantillon école est moins lourd en termes de charge de travail. En effet, selon cette procédure, un seul échantillon de 180 écoles est tiré avec une probabilité proportionnelle à la taille combinée des effectifs de 2^e année et de 6^e année du primaire. Ensuite, parmi ces 200 écoles, 100 écoles sont sélectionnées pour l'étude en deuxième. Cependant, comme les écoles sont sélectionnées sur base d'une taille combinée des effectifs de 2^e année et de 6^e année, certaines écoles ne proposant qu'un des deux niveaux peuvent être sélectionnées, réduisant ainsi la taille de l'échantillon de 2^e année ou de 6^e année. De plus, la taille combinée est une mesure moins précise pour un niveau déterminé et induit une plus grande variabilité des poids finaux des élèves. Cette variabilité accrue des pondérations est source d'instabilité dans les estimations des paramètres de population.

En 9^e/10^e année (fin de collège), le tirage de l'échantillon des élèves se fait dans 180 écoles. Le tirage de ces établissements est proportionnel à leurs effectifs. Cette procédure garantit la représentativité de l'échantillon au niveau national.

Pour des raisons de contraintes budgétaires, le PASEC a opté pour cette seconde alternative. Les procédures techniques d'échantillonnage expliquées ci-dessous s'y rapportent.

Chapitre 2

Échantillonnage : procédures techniques



La suite de ce chapitre est consacrée à la conception et à la réalisation du plan d'échantillonnage à trois degrés. Les procédures décrites sont celles directement applicables pour l'étude du PASEC2024 en vue de prélever un échantillon d'élèves de 9^e/10^e année au collège, un échantillon d'élèves de 6^e année et un échantillon d'élèves de 2^e année du primaire. Un échantillon principal de 180 écoles au collège et un échantillon de 200 écoles au primaire sont tirés de façon aléatoire proportionnellement aux effectifs des classes concernées (pour le primaire, il s'agit d'une combinée des deux grades, 2^e et 6^e année). L'échantillon complet du primaire constitue l'échantillon d'écoles pour la 6^e année tandis qu'un sous-échantillon aléatoire et simple de 100 écoles est tiré pour la 2^e année, à partir de l'échantillon principal d'écoles.

2.1. Étape1 : la liste des écoles de la population ou *Base de sondage*

Pour être en mesure de tirer les échantillons d'une évaluation internationale de systèmes éducatifs, il importe de disposer d'une *Base de sondage* distincte par pays participant.

La *Base de sondage* est constituée de la liste complète des unités d'échantillonnage de premier degré (en anglais, *Primary Sampling Unit* ou PSU) qui compte au moins un individu de la ou des populations cibles. Dans le cadre de l'enquête PASEC, la *Base de sondage* reprend l'ensemble des écoles fréquentées par au moins un élève de 2^e primaire et/ou un élève de 6^e primaire de la population cible.

Dès lors, il convient de préciser autant que faire se peut, ce qu'on entend par « école ». Il n'est pas aisé de définir, à un niveau international, le concept d'école et pour preuve, les manuels d'échantillonnage des études IEA (TIMSS, PIRLS...) et de l'OCDE (PISA, TALIS) pour ne citer que les plus importantes dans le domaine, se contentent d'éviter la question. Cette difficulté résulte essentiellement de la diversité des spécificités nationales et des différents niveaux de pouvoirs décisionnels. Ainsi, à titre d'exemple, en Communauté française de Belgique, un établissement scolaire peut se définir comme suit :

- une entité physique, dénommée *implantation*, et localisée en un seul lieu géographique bien circonscrit ;
- une unité administrative qui peut regrouper plusieurs sous-entités ou implantations localisées en différents lieux. Il s'agit par exemple, d'établissements primaires de plusieurs villages qui dépendent, pour des raisons purement administratives, d'un seul directeur.

Dans le cadre de l'étude PASEC dont la finalité principale est l'identification de variables *Maître* et *École* associées à une plus grande performance, il semble logique de proposer l'*implantation* comme définition de l'établissement scolaire, c'est-à-dire l'entité physique et non l'entité administrative. En effet, comme les variables écoles sont susceptibles de varier parfois substantiellement d'une entité physique à l'autre au sein d'une unité administrative, utiliser ce niveau comme définition de l'école nécessiterait théoriquement d'administrer autant de questionnaires école qu'il y a d'entités physiques. En d'autres termes, une variable école serait susceptible de varier au sein d'une école, ce qui est statistiquement un non-sens.

Si un pays ne peut fournir des informations valides au niveau *implantation*, l'entité administrative peut néanmoins être utilisée à défaut, comme unité principale dans ce pays. Quelle que soit l'option prise, TOUTES les écoles d'une *Base de sondage* nationale doivent répondre à la même définition.

Dans les pays concernés par l'enquête du PASEC, la *Base de sondage* sera établie à partir de l'annuaire statistique des ministères en charge de l'éducation du pays. Chaque responsable national doit envoyer au PASEC la liste d'écoles correspondant à sa population cible, c'est-à-dire après éventuelle réduction de couverture suite à des contraintes budgétaires ou politiques, par exemple. Il s'agit donc de l'ensemble des écoles qui comptent des élèves de 6^e année du primaire et/ou de 2^e année du primaire.

En plus de la *Base de sondage* nationale, les responsables nationaux doivent communiquer diverses informations s'y rapportant. Il s'agit en quelque sorte des clés de lecture de la *Base de sondage*. Le responsable national devra ainsi :

- **dater les données**, c'est-à-dire mentionner l'année académique couverte par la *Base de sondage*. Les Ministères en charge de l'éducation ne disposent généralement pas de données concernant l'année académique en cours mais au mieux les données de l'année précédente. Chaque pays doit fournir les données les plus récentes dont il dispose et indiquer l'année académique couverte. Notons que la qualité des données doit primer sur leur récence. On préférera les données vérifiées de 2023 aux données en cours de nettoyage de 2024 à titre d'exemple ;

- **documenter les exclusions : réduction de couverture et exclusion d'écoles.** Une réduction de la couverture de la population cible doit être documentée et les taux de population concernés doivent être calculés et communiqués à l'équipe du PASEC. Les causes d'exclusions d'écoles (écoles spécialisées, écoles « inaccessibles ») doivent également être renseignées. Ces exclusions d'écoles devront impérativement être limitées à 5 % pour ne pas mettre en péril la comparabilité internationale. Des formulaires ont été conçus à cette fin ;
- **définir les variables de stratification.** Le PASEC conseille mais n'exige pas l'utilisation de variables de stratification. De plus, une variable de stratification ne peut être utilisée que si cette variable est disponible pour toutes les écoles de la *Base de sondage*. Il est par exemple inutile d'envisager de stratifier un échantillon sur base du type d'écoles si on ne dispose pas de l'information pour chaque école ou si l'information dont on dispose est très peu fiable, voire même changer d'une année à l'autre comme c'est le cas avec les classes à simple ou à double vacation à titre d'exemple.

La *Base de sondage* nationale sera transmise au PASEC sous la forme d'un fichier Excel. Chaque ligne de ce fichier Excel représente une et une seule école. Les informations suivantes seront fournies pour chacune des écoles de la liste :

- **un identifiant unique.** Si un numéro d'identification national existe, celui-ci peut être utilisé. À défaut, un numéro d'identification unique devra être attribué à chacune, deux écoles différentes ne pouvant avoir le même identifiant. Nous conseillons de nommer la variable identifiant. En plus de l'identifiant, le nom de l'école, l'adresse, un numéro de téléphone s'il existe, et une personne de contact sont préconisés ;
- **la taille.** L'enquête du PASEC cible deux populations distinctes au primaire (une seule cible au collège – 9^e/10^e année), deux variables de taille sont donc attendues : le nombre d'élèves inscrits en 6^e année du primaire pour l'échantillon principal de l'étude et le nombre d'élèves inscrits en 2^e année du primaire pour l'échantillon de l'étude « secondaire ». Il se peut cependant que certains Ministères de l'éducation ne disposent pas de statistiques avec un tel niveau de précision mais seulement du nombre total d'élèves dans l'école. Le pays pourra alors fournir pour chaque école, le nombre total d'élèves ET le nombre de niveaux scolaires organisés dans l'école. La taille sera alors estimée en divisant le nombre total d'élèves par le nombre de niveaux scolaires organisés. Deux remarques importantes doivent encore être faites concernant la mesure de taille :
 - une école pour laquelle aucune information de taille n'est communiquée ne pourra être prise en compte pour tirer un échantillon PPS. Les centres nationaux devront donc mettre tout en œuvre pour fournir une mesure de taille pour chacune des écoles. Si, pour quelque raison que ce soit, il s'avère impossible d'associer à certaines écoles une mesure de taille, il est fortement conseillé de ne pas exclure ces écoles mais plutôt de les regrouper dans une strate explicite. Certaines seront ensuite sélectionnées selon une procédure aléatoire et simple, et ce, pour maintenir un taux de recouvrement maximal et ainsi garantir la représentativité de l'échantillon ;
 - le même type de mesure de taille doit être utilisé pour toutes les écoles d'un pays donné. La mesure de taille ne peut être le nombre d'élèves du niveau pour certaines écoles et pour d'autres, une approximation au départ de la population totale ;
- **les variables de stratification.** À nouveau, et au risque de se répéter, il est essentiel de disposer des informations pour chaque école reprise dans la *Base de sondage*. Si une variable de stratification est manquante, même pour une seule école de la liste, cette stratification ne pourra être appliquée puisque cette (ou ces) école(s). Il serait alors nécessaire, comme évoqué ci-dessus avec la taille des écoles, de regrouper ces écoles dans une strate explicite et de tirer les écoles indépendamment de cette variable. L'effet escompté de réduction de la variance d'échantillonnage s'en verrait affecté ;
- **les exclusions.** Les écoles fréquentées uniquement par des élèves qui répondent aux cotes d'exclusion sont incluses dans la *Base de sondage*, le statut d'exclusion et la raison justifiant celle-ci sont documentés.

Un extrait de *Base de sondage* est présenté dans la Figure 4 ci-dessous.

FIGURE 4 : Extrait d'une *Base de sondage* fictive

Identifiant unique		Coordonnées de l'école				Mesures de taille		Variables de stratification			Exclusions	
A	B	C	D	E	F	G	H	I	J	K	L	M
Identifiant école	Nom école	Localité	Adresse	Personne de contact	Tél	Nombre élèves - 6 ^e primaire	Nombre élèves - 2 ^e primaire	Région	Type d'école	Urbanisation	Exclusion n Oui/Non	Raison d'exclusion
15104						42	38	Nord	Publique	Rural	Non	
15128						75	84	Sud	Publique	Urbain	Non	
15129						5	7	Sud	Privée	Rural	Oui	Très petite école
15131						40	50	Nord	Privée	Rural	Non	
...												
15258						120	100	Est	Publique	Urbain	Non	
16301						25	22	Sud	Privée	Rural	Oui	Accès très difficile

Source : Auteurs

Au primaire, un responsable national pourrait légitimement vouloir proposer deux bases de sondage distinctes pour la 6^e année et la 2^e année. Ce procédé n'est cependant pas à encourager car le risque d'observer des incohérences entre les deux listes est grand. Par exemple, une même école pourrait avoir deux identifiants différents dans les deux bases de sondage, être attribuée à la strate d'écoles urbaines dans une liste et à la strate d'écoles rurales dans l'autre, etc.

Lorsqu'il dispose de la *Base de sondage* d'un pays, le PASEC doit **valider les choix méthodologiques** faits par le pays. Dans de nombreux cas, des contacts directs (appels téléphoniques ou courriels) avec le responsable national s'avèreront nécessaires pour clarifier, étayer et valider certains choix. La validation portera notamment sur :

- la pertinence et l'ampleur des réductions de couverture et des exclusions d'écoles si le pays désire en appliquer ;
- le choix des variables de stratifications explicite et implicite supplémentaires ou différentes de celles préconisées par le PASEC ;
- la nécessité de sur-échantillonner pour répondre à d'éventuels désidérats nationaux.

Une étape de **nettoyage de la Base de sondage** est alors programmée. Cette étape consiste en une série de vérifications de cette *Base de sondage* :

- vérifier que certaines données ne sont pas manquantes : identifiant école, mesure de taille, variables de stratification ;
- s'assurer qu'une même école n'est pas répertoriée plusieurs fois dans la liste. Il existe différentes techniques pour vérifier l'existence de doublons. La plus simple, à nos yeux, consiste à réaliser un tableau de fréquence sur la variable « identifiant », et ensuite de repérer, manuellement ou de manière automatisée, les identifiants qui apparaissent plus d'une fois ;
- estimer la vraisemblance des mesures de taille communiquées :
 - comparer la taille totale de la population élève à des données officielles fournies par ailleurs comme par exemple les données de population par âge ou les données relatives à des enquêtes antérieures. Les données publiées par l'Institut de Statistiques de l'UNESCO peuvent à cet égard constituer une base de comparaison utile. Si des différences de plus de 10% apparaissent, il convient de contacter les centres nationaux pour identifier les raisons de ces différences et, éventuellement, corriger les bases de sondage ;
 - s'interroger sur la vraisemblance de la mesure de taille de l'école lorsque celle-ci s'écarte fortement de la moyenne nationale, +/- 3 écart-types ;

- comparer les effectifs de deuxième et de sixième et repérer les écoles atypiques. Ce repérage peut également se réaliser par l'intermédiaire d'une régression linéaire (les effectifs de 6^e année sont prédits par les effectifs de 2^e année et les résidus sont ensuite calculés).

Parfois de nouveaux contacts avec l'équipe nationale PASEC seront nécessaires pour finaliser la *Base de sondage*.

2.2. Étape 2 : attribution d'une mesure de taille 'MOS' à chaque école

Le plan d'échantillonnage au primaire prévoit qu'un seul échantillon de 200 écoles est tiré avec une probabilité proportionnelle à la taille combinée des effectifs de 2^e année et de 6^e année du primaire. Ensuite, parmi ces 200 écoles sélectionnées pour la 6^e année, 100 écoles sont sélectionnées à nouveau pour l'étude en deuxième. Au collège, un échantillon de 180 établissements est tiré avec une probabilité proportionnelle à la taille des effectifs des élèves de 9^e/10^e année uniquement (Kish, 1965).

Au primaire, la taille combinée d'une école sera donc calculée en sommant le nombre d'élèves de 2^e année et le nombre d'élèves de 6^e année de l'école. Concrètement, la taille combinée n'est cependant pas une variable directement utilisable pour l'échantillonnage. En effet, les procédures d'échantillonnage requièrent une mesure de taille par niveau scolaire. Cette taille par niveau scolaire (SIZEGRAD) sera donc calculée en divisant la taille combinée par deux.

À ce stade, la liste des écoles renferme différentes mesures de taille. Si le pays fournit les informations de taille comme souhaitées, la liste des écoles contient les variables de mesures de taille suivantes :

- le nombre d'effectifs en 6^e année (SIZE6)⁴ ;
- le nombre d'effectifs en 2^e année (SIZE2) ;
- la taille combinée, égale à la somme des effectifs de 6^e année et de 2^e année (SIZECOMB) ;
- la taille estimée par grade, égale à la taille combinée divisée par 2 (SIZEGRAD).

Si le pays fournit uniquement des effectifs globaux pour chacune des écoles, la liste des écoles contient :

- la taille globale, égale au nombre total d'effectifs dans l'école (SIZEGLOB). Dans ce cas, le nombre de grades organisés dans l'école est également donné (NB_GRAD) ;
- la taille estimée par grade, égale à la taille globale divisée par le nombre de grades (SIZEGRAD).

C'est cette procédure qui est également utilisée pour le collège où l'effectif de l'établissement est rapporté à l'effectif des classes de fin de scolarité du secondaire (SIZE9).

La variable de taille SIZEGRAD correspond à la taille par grade calculée au départ des données de la *Base de sondage*. Dans le meilleur des cas, la *Base de sondage* contient les données relatives à l'année scolaire précédente, mais elles peuvent aussi dater de deux voire trois ans. Comme expliqué précédemment, les écarts entre la mesure de taille indiquée dans la *Base de sondage* et le nombre d'élèves réel au moment de l'évaluation génèrent une variabilité des poids finaux des élèves (Ardilly, 2006). Par ailleurs, ces écarts ont d'autant plus d'impact que l'effectif de l'école est petit. Ainsi, une école qui renseignerait 100 élèves en sixième dans la *Base de sondage* et qui au moment de l'évaluation en compterait 120, introduirait une faible variation de poids. Par contre, une école qui gagnerait également 10 élèves mais doublerait son effectif (et passerait par exemple de 10 à 28 élèves) entraînerait une forte modification des poids finaux des élèves. C'est pourquoi, afin de réduire les variations de poids finaux des élèves, la mesure de taille des petites écoles est artificiellement ramenée à la taille théorique de l'échantillon par classe. Ainsi, une nouvelle variable de mesure de taille est introduite : il s'agit de la variable **MOS (measure of size)**.

La mesure de taille **MOS** est égale à la taille estimée par grade (SIZEGRAD) pour autant que celle-ci soit supérieure ou égale à la taille théorique de l'échantillon par classe (en anglais *Target Cluster Size* ou TCS), c'est-à-dire 28. Dans le cas contraire, MOS est égale à 28. Mathématiquement, on peut écrire :

$$MOS = \max(SIZEGRAD, 28)$$

Un exemple est présenté dans le *Tableau 10*.

4 Entre parenthèses sont des propositions de noms à attribuer à ces variables pour éviter toute confusion. Ces noms de variables seront utilisés dans la suite du document.

TABLEAU 10 : Exemple fictif de *Base de sondage*

A	B	C	D	E	F	G	H	I	J	K
Identifiant école	Nom école	Localité	Adresse	Personne de contact	Tél	SIZE6	SIZE2	SIZECOMB	SIZEGRAD	MOS
15104						42	38	80	40	40
15128						75	84	159	79,5	79,5
15129						5	7	12	6	28
15129						40	50	90	45	45
...										
15258						120	100	220	110	110
16301						5	22	27	13,5	28
...										

Source : Auteurs

La variable MOS est la variable mesure de taille sur base de laquelle sera calculée la probabilité de sélection de l'école.

2.3. Étape 3 : calcul des taux d'exclusion d'écoles

Pour garantir la représentativité de l'échantillon, le taux d'exclusion ne doit pas dépasser le seuil des 5 % d'élèves par niveau scolaire. Ces exclusions peuvent survenir (i) pour des raisons de composition de l'école (l'école est uniquement fréquentée par des élèves répondant aux critères d'exclusion d'élèves, c'est le cas des écoles pour élèves handicapés) ou (ii) pour des raisons exceptionnelles et individuelles (école trop petite ou inaccessible). Cette seconde catégorie d'écoles doit rester sous le pourcentage d'élèves.

À partir de la *Base de sondage* nettoyée, l'équipe du PASEC doit chiffrer les exclusions. Ces exclusions d'écoles sont quantifiées en termes de pourcentage d'élèves concernés et non en termes de pourcentage d'écoles. Comme les standards sont établis sur base d'un niveau scolaire donné, il convient de rapporter les taux d'exclusion séparément pour la sixième et pour la deuxième et donc de calculer les exclusions à partir de la mesure de taille par grade (les variables SIZE2 et SIZE6). Si ces variables ne sont pas disponibles, les taux d'exclusion seront calculés en utilisant la variable SIZEGRAD.

Les pourcentages sont calculés par type d'exclusion. Si ceux-ci respectent le seuil établi, les écoles peuvent être définitivement exclues de la population effective. Dans le cas contraire, il faudra négocier avec le centre national pour réduire ces exclusions. Par exemple, si le taux d'élèves inscrits dans des écoles comptant moins de 8 élèves est inférieur à un pourcent, ces écoles sont retirées de la *Base de sondage*, sinon elles sont maintenues et ultérieurement intégrées à une strate de petites écoles (voir plus loin).

Il va de soi que les différents taux d'exclusion devront ensuite être soigneusement renseignés dans le rapport technique de l'étude.

Une fois les exclusions définitivement traitées, la *Base de sondage* ne contient plus que des écoles effectivement éligibles. La liste des écoles exclues (c'est-à-dire les lignes ou observations correspondantes) est conservée dans un fichier indépendant.

2.4. Étape 4 : le traitement des petites et très petites écoles

Pour les enquêtes du PASEC, une **très petite école** est définie comme une école qui compte moins de 8 élèves pour un niveau, soit SIZEGRAD plus petit que huit. Comme précédemment mentionné, ces très petites écoles seront exclues de l'échantillon pour autant que le nombre total d'élèves inscrits dans ces écoles ne dépasse pas 1 % de la population cible.

La procédure qui est décrite ci-dessous pour le primaire est également appliquée au collège pour un échantillon standard national de 180 établissements avec 30 élèves tirés aléatoirement dans la classe à évaluer dans l'école.

Une **petite école** est une école qui compte moins d'élèves au niveau cible que le nombre d'élèves à échantillonner dans l'école, c'est-à-dire que l'école compte moins de 28 élèves (SIZEGRAD < 28).

Par opposition, une « **grande** » école est une école fréquentée par au moins 28 élèves du niveau cible (SIZEGRAD ≥ 28).

Si le pourcentage d'élèves inscrits dans des petites écoles est égal ou supérieur à 4 %, une strate explicite de petites écoles est créée pour garantir une plus grande stabilité des poids. Au sein de cette strate, les écoles sont sélectionnées suivant une procédure strictement aléatoire et non par PPS. Comme la procédure d'échantillonnage des écoles prévoit une sélection selon la taille sur base de la variable MOS (voir plus loin), attribuer une valeur fixe de 28 pour le MOS des petites écoles revient à tirer un échantillon aléatoire et simple dans la strate *petites écoles*.

Si les très petites écoles ne peuvent être exclues car le pourcentage d'élèves qu'elles comptent est supérieur à 1 %, ces très petites écoles sont intégrées à la strate *petites écoles*.

Comme expliqué et illustré ci-après, le nombre d'écoles à échantillonner dans la strate de petites écoles et/ou de très petites écoles est proportionnel au nombre d'élèves qui fréquentent ces écoles et non proportionnel au nombre d'écoles dans la strate. Toutefois, si un ajustement n'est pas opéré, l'échantillon final comptera moins d'élèves qu'attendu.

Pour déterminer le nombre d'écoles à échantillonner dans la strate *petites écoles*, il convient de :

1. calculer le pourcentage d'élèves qui fréquentent cette strate, en utilisant la variable SIZEGRAD ;
2. multiplier ce pourcentage par la taille attendue de l'échantillon élèves de 6^e année – soit 200 écoles * 28 élèves, c'est-à-dire 5600 élèves – pour obtenir le nombre d'élèves à tester dans la strate *petites écoles* ;
3. calculer la taille moyenne des écoles dans la strate *petites écoles*, au départ de la variable SIZEGRAD ;
4. diviser le nombre d'élèves à tester dans la strate par la taille moyenne de ses écoles pour obtenir le nombre d'écoles à tester.

Par exemple, un pays compte 10% d'élèves fréquentant une petite école. Il est donc fortement conseillé de rassembler les écoles dans une strate explicite. L'échantillon global de 5600 élèves devra donc être composé de 560 élèves provenant de petites écoles, les 5040 autres élèves provenant de *grandes écoles* (soit 180 écoles et 28 élèves dans chacune). Pour déterminer le nombre d'écoles à tester, il convient de calculer la taille moyenne au sein de cette strate de *petites écoles*, disons 14, et de diviser le nombre d'élèves à tester dans cette strate, soit 560, par 14. Il sera nécessaire d'inclure 40 (560/14) petites écoles dans l'échantillon. La taille de l'échantillon final d'écoles sera égale à 220 (40 + 180), et non 200.

S'il importe, pour des raisons administratives, organisationnelles ou financières, de limiter la taille de l'échantillon écoles, il s'agira de procéder comme pour les autres strates, à savoir :

1. calculer le pourcentage d'élèves dans la strate *petites écoles* ;
2. multiplier ce pourcentage par 200 pour obtenir le nombre d'écoles à tester dans la strate.

Dans le cadre de notre exemple, il faudrait donc tester 20 écoles au lieu de 40 écoles. Comme la taille moyenne est de 14, cela signifie que 280 élèves devraient être testés, au lieu de 560.

2.5. Étape 5 : calcul de la taille de l'échantillon d'écoles pour chaque strate explicite

L'échantillon de 200 écoles (180 établissements pour le collège) est fragmenté en autant de sous-échantillons qu'il y a de strates explicites. Le nombre d'écoles à échantillonner par strate peut, soit être proportionnel à la taille de la strate (Neyman, 1934), soit fixé pour uniformiser les variances d'échantillonnage d'une strate à l'autre. Dans ce dernier cas, les échantillons seront, à peu de choses près, de taille identique d'une strate à l'autre. La proportionnalité est particulièrement recherchée lorsqu'il s'agit de produire des estimations pour l'ensemble de la population. Par contre, l'uniformisation des variances d'échantillonnage s'avère particulièrement efficace lorsqu'il s'agit principalement de comparer les résultats entre strates explicites.

Ainsi, comme mentionné précédemment, la variance d'échantillonnage pour l'estimation de la moyenne, dans le cadre d'un échantillon aléatoire simple, est égale à :

$$\sigma_{(\hat{\mu})}^2 = \frac{\sigma^2}{n}, \text{ avec } n \text{ représentant la taille de l'échantillon.}$$

Par ailleurs, si l'échantillon a été stratifié, on peut démontrer que la variance d'échantillonnage pour l'estimation de la moyenne s'écrit :

$$\sigma_{(\hat{\mu})}^2 = \frac{\sum_{h=1}^H N_h \sigma_{h(\hat{\mu})}^2}{N^2}, N_h \text{ étant la taille dans la strate}$$

Ceci revient à calculer la somme pondérée des variances d'échantillonnage de chaque strate, divisée par le carré de la taille de la population.

Enfin, toujours dans le cadre d'un échantillon aléatoire et simple, la variance d'échantillonnage sur la différence de moyennes entre deux strates explicites est égale à :

$$\sigma_{(\hat{\mu}_1 - \hat{\mu}_2)}^2 = \sigma_{(\hat{\mu}_1)}^2 + \sigma_{(\hat{\mu}_2)}^2$$

Le *Tableau 11* présente la taille de la population de deux strates, la taille de l'échantillon, la variance de population de la variable mesurée, la variance d'échantillonnage pour l'estimation de la moyenne globale et la variance d'échantillonnage pour l'estimation de la différence entre les deux strates.

TABLEAU 11 : Différences entre les variances d'échantillonnage selon la taille de l'échantillon par strate

Strate	Taille de la population	Taille de l'échantillon	Écart-type	Var. Éch moyenne strate	Var. Éch moyenne globale	Var. Éch. Différence
1	10000	1000	100	3,2	2,70	13,2
2	1000	100	100	10,0		
1	10000	550	100	4,3	3,56	8,5
2	1000	550	100	4,3		

Source : Auteurs

Si 1000 et 100 sujets sont respectivement prélevés des strates 1 et 2 (soit 10 % de la population), la variance d'échantillonnage de la moyenne générale sera de 2.70 et la variance d'échantillonnage de la différence de moyenne entre les deux strates sera égale à 13,2. Par contre, si 550 sujets sont prélevés au sein de chaque strate, la variance d'échantillonnage pour la moyenne générale augmente. Par contre, elle diminue pour la différence de moyennes entre les deux strates.

Dans le cadre des enquêtes PASEC, la production d'indicateurs nationaux est prioritaire par rapport aux comparaisons infranationales. La taille des échantillons sera donc généralement proportionnelle à la taille des states.

Toutefois, convient-il d'échantillonner un nombre d'écoles proportionnel au nombre d'écoles dans la strate ou au contraire échantillonner un nombre d'écoles proportionnel au nombre d'élèves dans la strate ?

Imaginons une *Base de sondage* fictive composée de 2 strates explicites comprenant chacune 5 écoles. Le plan d'échantillonnage stipule un échantillonnage de 4 écoles et ensuite de 10 élèves dans chacune d'elles, soit un échantillon final de 40 élèves (voir *Tableau 12* ci-dessous).

TABLEAU 12 : Base de sondage fictive : 2 strates de 5 écoles de tailles différentes

	Strate	Identifiant école	Taille
5 écoles 100 élèves	1	1	20
	1	2	20
	1	3	20
	1	4	20
	1	5	20
5 écoles 300 élèves	2	6	60
	2	7	60
	2	8	60
	2	9	60
	2	10	60

Source : Auteurs

Envisageons les deux possibilités.

1^{re} option : le nombre d'écoles échantillonnées par strate est calculé en proportion du nombre d'écoles dans la strate. Les proportions d'écoles dans la strate 1 et dans la strate 2 sont semblables et égales à 50 % (5/10). On sélectionnera donc 2 écoles dans chaque strate. Le *Tableau 13* montre que selon cette procédure, le poids final des élèves sera sensiblement différent d'une strate à l'autre.

TABLEAU 13 : Échantillonnage d'écoles proportionnel au nombre d'écoles dans la strate

Strate	Identifiant école	Taille	Poids école w_i	Poids élève dans l'école w_{ij}	Poids final de l'élève w_{ij}
1	1	20			
1	2	20			
1	3	20	2.5	2	5
1	4	20			
1	5	20	2.5	2	5
2	6	60			
2	7	60	2.5	6	15
2	8	60			
2	9	60	2.5	6	15
2	10	60			

Source : Auteurs

2^e option : le nombre d'écoles échantillonnées par strate est calculé en proportion du nombre d'élèves dans la strate. La strate 1 inclut 25 % des élèves (100/400) tandis que 75 % des élèves sont dans la strate 2. Par cette procédure, la stabilité des poids finaux des élèves est maintenue d'une strate à l'autre (Tableau 14).

TABLEAU 14 : Échantillonnage d'écoles proportionnel au nombre d'élèves dans la strate

Strate	Nb d'écoles	Nb d'élèves	Proportion d'élèves	Nb d'écoles à échantillonner	w_i	w_{ij}	w_{ji}
1	5	100	25%	1	5	2	10
2	5	300	75%	3	5/3	6	10

Source : Auteurs

Par cette simple simulation, il apparaît clairement que le nombre d'écoles à sélectionner au sein de chacune des strates explicites de la *Base de sondage* sera déterminé en fonction du nombre d'élèves inclus dans la strate sans tenir compte du nombre d'écoles. En effet, seule cette procédure permet de garantir l'équiprobabilité. Notons toutefois que cet exemple est fictif, puisqu'il est nécessaire d'échantillonner au moins quatre écoles par strate explicite, pour en garder au minimum deux pour le grade 2, pour être en mesure d'estimer la variance d'échantillonnage par des méthodes de réplication.

Pratiquement, le nombre réel d'élèves dans la strate h sera calculé en sommant les tailles estimées par grade de toutes les écoles de la strate comme suit :

$$N_{1_h} = \sum_{i=1}^{N_{sc_h}} SIZEGRAD_i'$$

tandis que le nombre d'écoles à échantillonner dans la strate n_{sc_h} est égal à :

$$n_{sc_h} = n_{sc} * \frac{N_{1_h}}{N} \text{ avec } n_{sc} \text{ égal à } 200.$$

2.6. Étape 6 : sélection des 200 écoles primaires de l'échantillon principal

Au sein de chaque strate explicite, les écoles sont sélectionnées avec une probabilité proportionnelle à leur taille. Pour ce faire, les écoles sont triées selon la variable **MOS** au sein de chaque strate explicite, et ensuite les écoles sont sélectionnées suivant une procédure systématique. Ainsi, l'échantillon aura une meilleure représentativité puisqu'il comportera à la fois de très grandes écoles, des écoles de taille moyenne et de écoles plus petites. Sans ce tri et sans l'application de cette procédure systématique, l'échantillon pourrait ne comporter que des grandes écoles, voire que des écoles assez petites.

Si des variables de stratification implicite sont utilisées, la liste d'écoles doit être triée selon ces variables également. Il convient alors d'alterner un tri ascendant et un tri descendant des observations. À titre d'illustration, imaginons que le pays propose d'utiliser deux variables de stratification SI1 et SI2, comportant chacune 3 modalités. Le *Tableau 15* présente comment les données devraient être ordonnées au sein de chaque strate explicite.

TABLEAU 15 : Exemple de tris successifs selon les variables de stratification implicite

Stratification implicite S1	Stratification implicite 2	MOS
1	1	Ascendant
1	2	Descendant
1	3	A
2	3	D
2	2	A
2	1	D
3	1	A
3	2	D
3	3	A

Source : Auteurs

Les observations sont d'abord triées selon la variable S1, puis selon la variable S2 et enfin selon la taille MOS. Pour la modalité 1 de la variable S1, les modalités de la variable S2 sont triées selon un ordre ascendant. Par contre, pour la modalité 2 de S2, elles sont triées selon un ordre descendant, et enfin selon un ordre ascendant pour la troisième modalité de S2. Grâce à ce procédé, on maximalise la ressemblance entre deux observations adjacentes inter et intra strates implicites, ce qui est particulièrement important pour le choix des écoles de remplacement.

Ce tri réalisé, il convient de numéroté les écoles de 1 à N_{sc} .

L'échantillonnage systématique est alors implémenté comme suit :

- Numéroté les i élèves de la liste d'écoles triée de 1 à N_{2_h} avec N_{2_h} le nombre d'élèves dans la strate

$$E_i = i \text{ avec } i = 1, \dots, N_{2_h}$$

Cette numérotation des élèves doit être réalisée sur base de la variable MOS.

- Calculer le pas d'échantillonnage en divisant le nombre total d'élèves dans la strate par le nombre d'écoles à échantillonner dans celle-ci :

$$Int = \frac{N_{2_h}}{n_{sc_h}}$$

Avec le nombre total fictif d'élèves dans la strate N_{2_h} calculé en sommant les MOS de la strate comme suit :

$$N_{2_h} = \sum_{i=1}^{N_{sc_h}} MOS_i$$

- Tirer un nombre aléatoire a issu d'une distribution uniforme $[0;1]$. C'est cette étape qui confère le caractère aléatoire à l'échantillon d'écoles.

$$a \in [0; 1]$$

- Multiplier cette constante a par le pas d'échantillonnage et arrondir le nombre à l'unité supérieure. Le produit obtenu permettra d'identifier la première école à sélectionner. Il s'agit en quelque sorte d'une première estampille de sélection.

$$Est_1 = a * Int$$

L'école qui contient l'élève $E_i = Est_1$ sera sélectionnée.

- Ajouter le pas d'échantillonnage pour définir la deuxième estampille de sélection.

$$Est_2 = Est_1 + Int$$

- Répéter cette procédure systématique jusqu'à ce que le nombre d'écoles à échantillonner soit atteint.

Le *Tableau 16* illustre cette procédure à partir d'une strate de 10 écoles et 400 élèves. Quatre écoles sont sélectionnées.

TABLEAU 16 : Sélection PPS systématique de 4 écoles dans une strate de 400 élèves

Identifiant école	Taille de l'école	Numéros E_i des élèves		École échant.	
		De	à		
1	10	1	10	0	$Est_1 = 0.232 * 100 = 23.2$ $Est_2 = 23.2 + 100 = 123.2$ $Est_3 = 123.2 + 100 = 223.2$ $Est_4 = 223.2 + 100 = 323.2$
2	15	11	25	1	
3	20	26	45	0	
4	25	46	70	0	
5	30	71	100	0	
6	35	101	135	1	
7	40	136	175	0	
8	45	176	220	0	
9	80	221	300	1	
10	100	301	400	1	

Source : Auteurs

Dans cet exemple, la première estampille est égale à 23.2 (l'arrondi supérieur est 24). L'école 2 sera sélectionnée car elle contient l'élève 24. Seront également sélectionnées, les écoles 6, 9 et 10 puisqu'elles contiennent respectivement les élèves 124, 224 et 324.

2.7. Étape 6 bis : traitement des très grandes écoles ou « écoles certaines »

Une **très grande école** est une école dont la mesure de taille est supérieure à Int , le pas d'échantillonnage calculé. Théoriquement, la probabilité de sélection de telles écoles est supérieure à 1. Si elles sont maintenues dans la *Base de sondage*, elles seront sélectionnées plus d'une fois par la procédure systématique. Pour éviter les difficultés statistiques ou administratives liées à cette situation, ces écoles, par ailleurs dénommées « **écoles certaines** » sont assignées à une strate distincte et sont automatiquement sélectionnées dans l'échantillon. Il s'ensuit que leur probabilité de sélection est ramenée à 1.

L'identification des *écoles certaines* a lieu immédiatement après le calcul du pas d'échantillonnage et leur traitement doit être implémenté avant la mise en œuvre de la procédure systématique de sélection des écoles (étape 6).

Cette procédure est réalisée à l'intérieur de chaque strate explicite initiale puisque les pas d'échantillonnage varient d'une strate à l'autre. Dans la strate d'*écoles certaines*, les probabilités de sélection des écoles seront ramenées à 1, et par conséquent le poids école w_i sera aussi égal à 1.

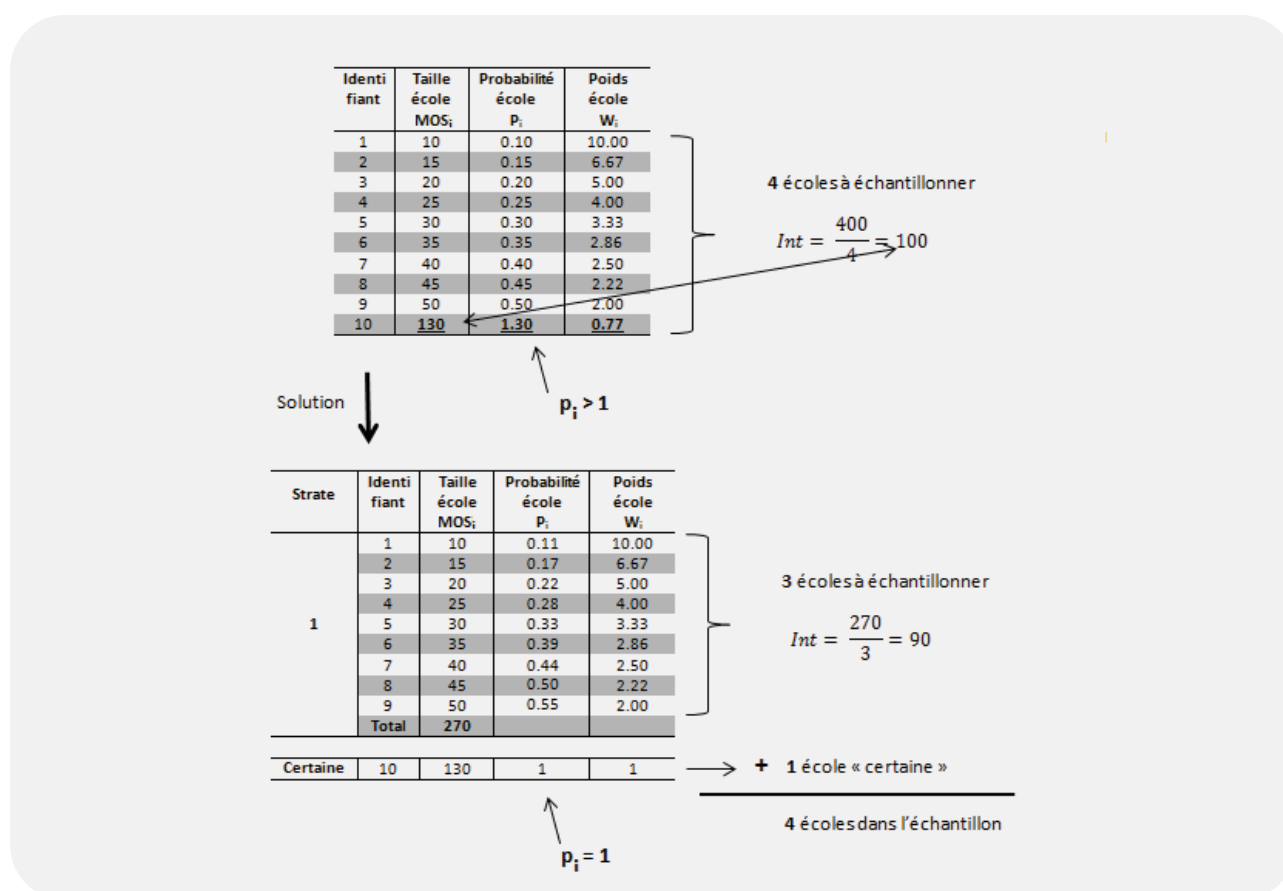
Repartons de l'exemple précédent, une strate de 10 écoles et 400 élèves, mais intégrons une très grande école dont la probabilité de sélection est supérieure à 1.

Dans la *Figure 5*, l'école 10, avec ses 130 élèves, a une probabilité de sélection égale à 1.30. Cette probabilité supérieure à 1 résulte du fait que le nombre d'élèves est supérieur au pas d'échantillonnage. Deux estampilles successives pourraient donc – selon le nombre aléatoire a – tomber à deux reprises sur des élèves appartenant à l'école 10, qui serait alors théoriquement sélectionnée deux fois.

Pour éviter ce non-sens, l'école 10 est placée dans une nouvelle strate appelée la « strate certaine ». La strate principale est maintenant constituée de 270 élèves répartis dans 9 écoles parmi lesquelles 3 doivent être sélectionnées. Le nouveau pas d'échantillonnage Int est égal à 90. Ce nouveau pas est supérieur à la taille maximale des écoles de la strate, ce qui revient à dire que toutes les écoles ont une probabilité de sélection inférieure à 1. Dans le cas contraire, les nouvelles *écoles certaines* devraient à leur tour être sorties de la strate principale et le pas d'échantillonnage devrait être à nouveau recalculé. L'échantillon final comprend donc 3 écoles sélectionnées dans la strate principale plus l'école certaine. Notons à nouveau que cet exemple fictif doit être ajusté à l'enquête PASEC puisque, dans ce cas, il est nécessaire d'avoir au minimum 4 écoles **non certaines** échantillonnées par strate explicite. Si un tel cas se présente, il est conseillé de fusionner cette strate explicite avec une autre strate explicite pour éviter une augmentation de l'échantillon *Écoles*.

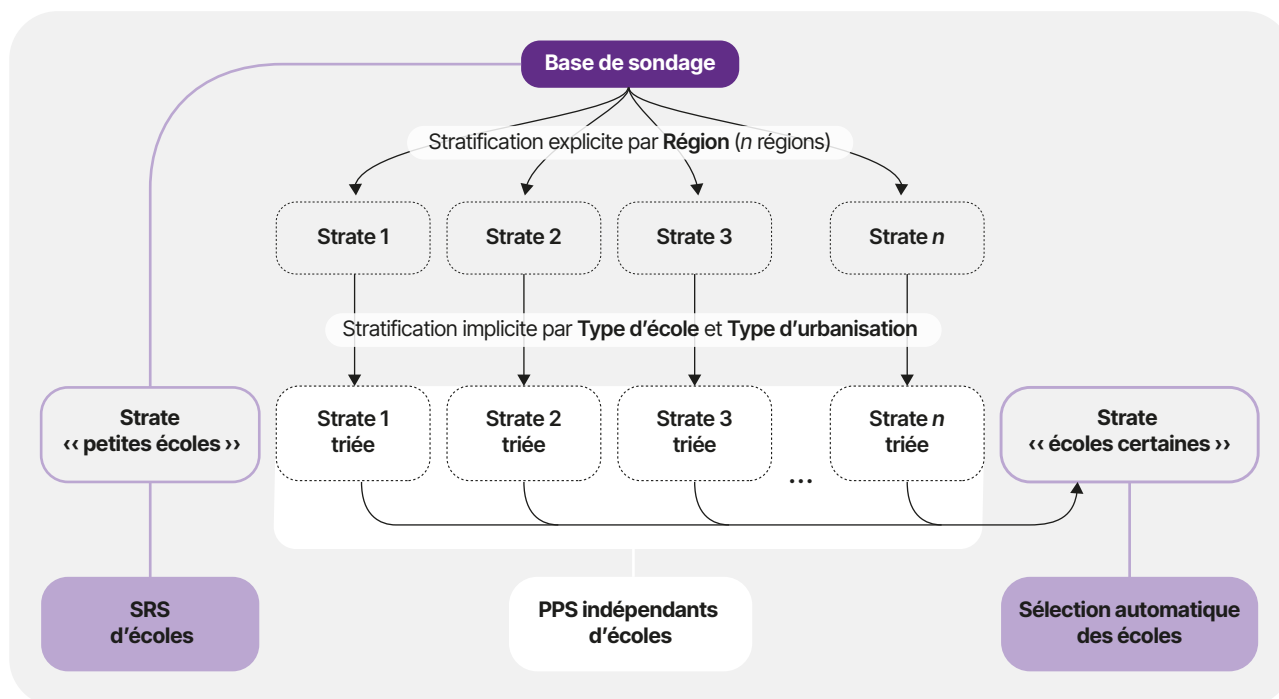
La Figure 6 ci-contre résume le processus d'échantillonnage des écoles en se basant sur les variables de stratification conseillées par le PASEC.

FIGURE 5 : Strate de 10 écoles et 400 élèves, échantillonnage de 4 écoles – 1 école certaine



Source : Auteurs

FIGURE 6 : Illustration du processus d'échantillonnage des écoles



Soit une *Base de sondage* nationale nettoyée comprenant les écoles qui organisent la 6^e primaire et/ou la 2^e primaire. La *Base de sondage* est stratifiée explicitement par région. Les n strates explicites sont triées selon deux variables de stratification implicite : le type d'école (publique/privée) et la variable d'urbanisation (urbain/rural). Si ce pays compte plus de 4 % d'élèves par grade inscrits dans une petite école (moins de 28 élèves du niveau), une strate *petites écoles* est créée. Les très grandes écoles, dont la probabilité de sélection au sein de la strate est supérieure à 1, sont sorties de leur strate initiale et regroupées dans une strate *écoles certaines*. La sélection des écoles se fait par strate : un échantillon SRS d'écoles dans la strate *petites écoles*, une sélection automatique dans la strate *écoles certaines* et n échantillons PPS d'écoles dans les n strates de *grandes écoles*.

2.8. Étape 7 : désignation des écoles de remplacement

Dans les enquêtes TIMSS, PIRLS ou encore PISA, il arrive que des écoles refusent de participer malgré les moyens et les efforts déployés pour éviter cette non-participation. Ce phénomène, assez variable d'un pays à l'autre, est extrêmement rare dans les pays membres de la CONFEMEN.

D'autres écoles échantillonnées pourraient ne pas pouvoir participer à l'évaluation pour cause de force majeure. Une école réquisitionnée comme centre électoral le jour de l'épreuve, une école momentanément fermée pour cause d'épidémie parmi les élèves, une école située dans un village inaccessible à l'administrateur de test suite à des inondations... sont des exemples de situations ne permettant pas la mise en place de l'évaluation.

Pour éviter une réduction de taille de l'échantillon, les écoles non-participantes pour situation exceptionnelle ou suite à un refus sont remplacées par une école dite « de remplacement ». Pour respecter le plan d'échantillonnage, une école de remplacement doit être « semblable » à l'école initialement sélectionnée, c'est-à-dire qu'elle doit faire partie de la même strate explicite et doit être de taille similaire. Un processus a été instauré pour identifier a priori deux écoles de remplacement potentielles à chacune des écoles de l'échantillon. Dans la *Base de sondage* ordonnée, et au sein de la strate, l'école qui suit directement l'école échantillonnée est désignée comme sa première école de remplacement (R1), l'école qui la précède est sa seconde école de remplacement (R2). En cas de non-participation de l'école initialement sélectionnée (S), c'est l'école R1 qui sera évaluée. L'école R2 est au besoin, l'école de remplacement de R1 dans le cas extrême d'un second refus.

Dans l'exemple décrit dans le *Tableau 17*, l'école 2 fait partie de l'échantillon. L'école 3 sera sa première école de remplacement et l'école 1 sera sa seconde école de remplacement.

Une école peut cependant n'avoir qu'une ou même aucune école de remplacement désignée. C'est le cas lorsque deux écoles sélectionnées se suivent directement dans la *Base de sondage* ou encore lorsqu'une école sélectionnée est en bordure de strate explicite (première ou dernière). Dans l'exemple ci-dessus, l'école 8 fait partie de l'échantillon et n'a qu'une seule école de remplacement. En effet, une même école ne peut avoir qu'un seul statut : l'école 7 est la première école de remplacement de l'école 6, elle ne peut être désignée comme seconde école de remplacement de l'école 8. L'école 10 n'a aucune école de remplacement puisqu'elle est la dernière école de la strate et que l'école qui la précède est déjà désignée comme école de remplacement de l'école 8.

TABEAU 17 : Désignation des écoles de remplacement R1 et R2

Identifiant école	MOS	École échant.	Statut de l'école
1	10	0	R2
2	15	1	S
3	20	0	R1
4	25	0	
5	30	0	R2
6	35	1	S
7	40	0	R1
8	45	1	S
9	80	0	R1
10	100	1	S

Source : Auteurs

Dans certains *surveys* comme PISA, une école peut théoriquement servir d'écoles de remplacement pour deux écoles de l'échantillon original et, selon les circonstances, servir d'écoles de remplacement pour l'une OU pour l'autre école. Par exemple, si l'école 10 refuse de participer et que l'école 8 accepte de participer, on pourrait utiliser l'école 9 comme école de remplacement de l'école 10.

2.9. Étape 8 : sélection des écoles de l'échantillon école de 2^e primaire

Le sous-échantillon de 100 écoles pour le test en 2^e primaire est constitué en sélectionnant au sein de chaque strate explicite et de façon aléatoire, la moitié des écoles déjà sélectionnées pour l'échantillon principal. À cette fin, il suffit de classer les écoles sélectionnées pour l'échantillon de 6^e selon un ordre purement aléatoire à l'intérieur de chaque strate, et ensuite de prendre les $n/2$ premières écoles pour l'échantillon de 2^e. Si un nombre impair d'écoles a été sélectionné pour l'échantillon de 6^e, on prendra soit les $(n+1)/2$ premières écoles, soit les $(n-1)/2$ premières écoles. Le choix entre $n+1$ ou $n-1$ se fera également sur une base aléatoire. On peut également décider de choisir $n+1$ pour les petites strates et $n-1$ pour les grandes strates pour assurer un minimum d'écoles par strate. Enfin, comme il a été déjà mentionné, au sein de chaque strate, il faut un minimum de 2 écoles sélectionnées et ce, afin de pouvoir estimer la variance d'échantillonnage.

Il se peut qu'une école sélectionnée ne compte pas d'élèves de sixième ou de deuxième. Dans ce cas, il ne faut surtout pas recourir à une école de remplacement. En effet, en procédant de la sorte, la taille de la population de sixième ou de deuxième serait surestimée. Par contre, une école inaccessible pour cas de force majeure le jour de l'évaluation, ou une école qui refuse de participer pourrait être remplacée.

2.10. Étape 9 : attribution d'un identifiant PASEC

Une fois les deux échantillons d'écoles formés, un identifiant PASEC (ID_PASEC) propre à l'étude est attribué à chaque école de l'échantillon principal ainsi qu'à chacune des écoles de remplacement. Cet identifiant a pour but de garder trace des écoles échantillonnées par strate explicite et par niveau scolaire, et de leurs écoles de remplacement respectives. Le code d'attribution est proche de celui utilisé dans l'enquête PISA. Il s'agit d'un identifiant unique à six chiffres : deux chiffres correspondant à la strate explicite, un chiffre correspondant au niveau(x) scolaire(s) évalués et trois chiffres d'identification de l'école :

- deux chiffres pour identifier la strate : les strates explicites sont numérotées séquentiellement 01, 02, 03, etc. ;
- un chiffre pour renseigner du ou des grade(s) évalué(s) : '1' si l'école est sélectionnée pour la 6^e uniquement, '2' si elle fait partie des deux échantillons ;
- trois chiffres d'identification de l'école :
 - au sein de chaque strate, les écoles sélectionnées (écoles 'S') pour l'échantillon principal (6^e primaire) sont numérotées séquentiellement à partir de 001. Par exemple, si 80 écoles sont sélectionnées dans une strate, les écoles 'S' sont numérotées de 001 à 080 ;
 - les trois chiffres d'identification des premières écoles de remplacement (R1) sont liés à l'identifiant de l'école sélectionnée 'S' correspondante. Il s'agit de l'identifiant de l'école 'S' augmenté de 300. Par exemple, la première école de remplacement de l'école échantillonnée 056 aura l'identifiant 356 ;
 - pour la seconde école de remplacement (R2), l'identifiant de l'école échantillonnée est augmenté de 600. Dans notre exemple, la seconde école de remplacement de l'école 056 recevra l'identifiant 656.

TABLEAU 18 : Exemple d'attribution des identifiants PASEC des écoles

Strate explicite	Identifiant école ID_NAT	Échantillon principal (6 ^e)	Statut de l'école (6 ^e) STATUT6	Échantillon secondaire (2 ^e)	Statut de l'école (2 ^e) STATUT2	Identifiant PASEC ID_PASEC
1	1	0	R2	0	R2	012601
1	2	1	S	1	S	012001
1	3	0	R1	0	R1	012301
1	4	0		0		
1	5	0	R2	0		011602
1	6	1	S	0		011002
1	7	0	R1	0		011302
1	8	1	S	1	S	012003
1	9	0	R1	0	R1	012303
1	10	1	S	0		011004
2	11	0		0		
2	12	0	R2	0	R2	022601
2	13	1	S	1	S	022001
2	14	0	R1	0	R1	022301
2	15	0		0		
2	16	0		0		
2	17	0	R2	0		021602
2	18	1	S	0		021002

Le *Tableau 18* ci-dessus illustre cette procédure et montre que l'identifiant PASEC est bien un identifiant unique. La compréhension du code d'identification aide à identifier les caractéristiques d'échantillonnage de l'école. Par exemple, l'école dont l'identifiant PASEC est 012003 est la troisième école échantillonnée dans la première strate de l'échantillon principal. Elle est sélectionnée pour les deux grades. L'école 021602 est la seconde école de remplacement (R2) de la deuxième école échantillonnée dans la strate 2. L'école 021002 à laquelle elle est liée ne fait pas partie du sous-échantillon de 2^e primaire. Dans notre exemple, l'école 021002 n'a pas de première école de remplacement R1, l'école 021602 sera donc l'école qui remplacera l'école 021002 si celle-ci devait l'être.

2.11. Finalisation de la liste des écoles échantillonnées

La liste des écoles échantillonnées, ou fichier *Échantillon*, contient les 180 écoles échantillonnées pour l'évaluation principale (ou un peu plus en raison des petites écoles) et leurs écoles de remplacement respectives, les autres écoles étant éliminées. Dans l'exemple repris dans le *Tableau 18*, les écoles 4, 11, 15 et 16 ne font pas partie de la liste.

Cette liste doit être envoyée dans une forme standardisée au responsable national. Si le PASEC opte pour l'envoi d'un tableau Excel, il est important de vérifier que l'exportation des données s'est opérée correctement. L'exportation de variables numériques en Excel pose parfois souci : vérifiez que les identifiants ont conservé leur longueur exacte et particulièrement ceux commençant par un zéro. D'autre part, les manipulations de données en Excel conduisent facilement à des transformations non conscientes de la base de données initiale. Il est donc fortement conseillé de fournir aux centres nationaux, une liste d'écoles échantillonnées dans un second format plus sécurisé ou une copie 'pdf' du fichier Excel.

Le fichier est ordonné de telle sorte qu'une école échantillonnée et ses écoles de remplacement sont groupées : une école échantillonnée est directement suivie par sa première et sa seconde école de remplacement.

Le fichier *Échantillon* doit au moins contenir les variables suivantes :

- le statut de l'école pour l'échantillon de 6^e année du primaire (STATUT6) ;
- le statut pour l'échantillon de 2^e année du primaire (STATUT2) ;
- l'identifiant PASEC (ID_PASEC) ;
- l'identifiant national de l'école (ID_NAT) ;
- le nom de l'école (NOM) ;
- la mesure de taille pour la 6^e année du primaire (SIZE6) ou la mesure de taille moyenne par grade (SIZEGRAD) si SIZE6 n'est pas disponible pour le pays. Si l'école est échantillonnée pour la 2^e année du primaire, la variable SIZE2 est également fournie ;
- la mesure de taille pour la 2^e année du primaire (SIZE2). La valeur de la variable n'est donnée que pour les écoles sélectionnées pour la 2^e année du primaire également (et ses écoles de remplacement). Si l'école ne fait pas partie de l'échantillon de 2^e, la valeur pour cette variable est laissée manquante ;
- la strate explicite (ID_STRATE) ;
- la mesure de taille qui a été utilisée pour le tirage (MOS) ;
- les probabilités de sélection pour chacune des années d'études (PROB6 et PROB2).

Le *Tableau 19* montre le format conseillé d'un fichier *Échantillon*.

TABLEAU 19 : Fichier *Échantillon* (exemple pour un pays qui dispose des tailles pour chaque grade)

	NOM	ID_ PASEC	ID_ ID_NAT	ID_ STRATE	STATUT6	STATUT2	SIZE6	SIZE2	PROB6	PROB2	MOS
1	XXX	012001	20EY07	01	S	S	21	24			21
2	XXX	012301	16AW00	01	R1	R1	21	26			21
3	XXX	012601	06HF02	01	R2	R2	19	23			20
4	XXX	011002	19XH04	01	S		28				28
5	XXX	011302	24RW03	01	R1		30				30
6	XXX	011602	00XA01	01	R2		29				29
7	XXX	012003	02VN01	01	S	S	56	60			56
8	XXX	012303	16BK00	01	R1	R1	58	59			58
9	XXX	011004	02ME00	01	S		85				85
10	XXX	022001	25EF02	02	S	S	35	40			35
11	XXX	022301	02EX00	02	R1	R1	37	39			37
12	XXX	022601	25GA01	02	R2	R2	35	38			35
13	XXX	021002	04IX01	02	S		62				62
14	XXX	021602	20DF01	02	R2		63				63

2.12. Vérification du fichier *Échantillon*

Avant de transmettre l'échantillon aux pays, il importe de vérifier si aucune erreur n'a été introduite dans la procédure. En effet, si des procédures standards doivent être développées afin d'automatiser au maximum le processus d'échantillonnage des écoles, des adaptations seront toujours nécessaires, ne serait-ce pour que les données nationales s'ajustent au format international, ou pour rendre compatible les variables de stratifications proposées par les centres nationaux.

Ci-dessous figure une liste non-exhaustive de vérification qu'il importe d'implémenter avant de renvoyer l'échantillon aux centres nationaux.

1. Au départ de la *Base de sondage*, après suppression dans le fichier informatique des écoles exclues, et après avoir, si nécessaire, regrouper les petites et très petites écoles dans une strate spécifique et les écoles certaines dans une autre strate, calculer par strate explicite :
 - le nombre total d'écoles ;
 - le nombre total d'élèves en sixième et en deuxième ;
 - la somme des *Measure of Size* (MOS) ;
 - le pas d'échantillonnage, c'est-à-dire la somme des MOS, divisé par le nombre d'écoles à échantillonner.
2. Au départ du fichier *Échantillon*, et qui sera transmis aux centres nationaux, calculer, par strate explicite :
 - vérifier que chaque école échantillonnée ou de remplacement a bien un identifiant unique, a une probabilité de sélection comprise entre 0 et 1 et ne présente aucune donnée manquante pour les différentes variables utilisées dans le processus d'échantillonnage ;
 - vérifier que chaque école, en dehors de la strate des écoles certaines, a une MOS supérieure au pas d'échantillonnage ;
 - le nombre d'écoles échantillonnées et comparer avec les nombres obtenus précédemment (voir section 2.5 Étape 5) ;
 - le nombre d'écoles de remplacement 1 et de remplacement 2 et vérifier que leur nombre ne dépasse pas le nombre d'écoles échantillonnées ;
 - pour les seules écoles échantillonnées, calculer la somme des probabilités de sélection. Cette somme devrait être égale au nombre d'écoles échantillonnées au sein de la strate ;

- pour les seules écoles échantillonnées, calculer le poids des écoles, à savoir $w_i = \frac{1}{p_i}$. Pondérer les données avec cette variable poids et calculer la somme des MOS par strate. Cette somme devrait être égale à la somme des MOS calculée sur l'ensemble des écoles ;
- pour toute école échantillonnée qui ne dispose pas à la fois d'une école de remplacement 1 et d'une école de remplacement 2, vérifier que ce résultat est dû à un des cas de figure présenté précédemment, à savoir l'école est au début ou à la fin de la liste d'écoles, ou les écoles sont de grandes tailles, comparativement au pas d'échantillonnage et il est impossible d'associer à chaque école échantillonnée une école de remplacement.



1

2

3

4

5

Chapitre 3

Échantillonnage des classes et des élèves



3.1. Sélectionner la classe et les élèves

Dans le cadre des études internationales menées sous l'égide de l'IEA ou de l'OCDE, l'échantillonnage des classes et/ ou des élèves relève de la responsabilité des centres nationaux. Ces derniers doivent prendre contact avec les directions ou leurs administrations pour obtenir la liste des élèves et ou des classes, pour ensuite tirer, école par école, l'échantillon et finalement, le communiquer aux écoles.

Dans le cadre des études du PASEC, il a été jugé préférable, pour diverses raisons, de ne pas suivre cette procédure. Dès lors, l'échantillonnage des classes et des élèves d'une école sélectionnée est confié aux administrateurs de test lorsque ceux-ci se rendent dans l'école le jour de l'évaluation. La procédure à suivre est identique au primaire pour la 6^e et la 2^e année et au collège en classe de 9^e/10^e année.

Pour échantillonner la classe du niveau cible et les élèves de cette classe qui prendront part à l'évaluation, l'administrateur de test doit procéder comme suit :

- noter le nombre de classes du niveau cible organisé dans cette école. Si l'école fonctionne en double vacation, la classe du matin et la classe de l'après-midi doivent être considérées comme deux classes distinctes ;
- si plusieurs classes sont organisées, tirer au sort la classe qui sera sélectionnée. À cette fin, il suffit de préparer autant de vignettes qu'il y a de classes, de les placer dans un chapeau et d'en tirer une au hasard ;
- si l'école fonctionne en double vacation, la classe automatiquement sélectionnée sera celle organisée au moment de l'évaluation. Il s'ensuit que l'administrateur de test devra adapter son horaire d'évaluation à celui de la classe échantillonnée : si le second jour d'évaluation, l'horaire des classes est inversé, l'administrateur de test devra s'organiser pour tester les mêmes élèves. Cette procédure ne biaise pas l'échantillon pour autant que la constitution des classes en double vacation c'est-à-dire la répartition des élèves aux seins de celles-ci repose sur un processus aléatoire ou quasi aléatoire. Dans le cas contraire, si une répartition non aléatoire des élèves est suspectée, il faudra nécessairement procéder à un tirage aléatoire entre la classe du matin et celle de l'après-midi. À titre d'exemple, la classe devra être échantillonnée par tirage aléatoire si les élèves les plus faibles sont préférentiellement affectés à une des deux classes ou encore si les garçons sont majoritairement affectés à une classe et les filles à l'autre ;
- dans la classe sélectionnée, tirer l'échantillon d'élèves :
 - si la classe comprend 28⁵ élèves ou moins (élèves absents compris), tous les élèves de la classe sont sélectionnés. Parmi ceux-ci, seuls les élèves présents passeront l'évaluation. Les élèves absents sont assimilés à des non-réponses. Le nombre d'élèves absents doit impérativement être répertorié afin qu'ils puissent être pris en considération lors de la pondération des données ;
 - si la classe comprend plus de 28 (élèves absents compris), un échantillon aléatoire et simple d'élèves doit être constitué. Pour ce faire, l'administrateur de test peut utiliser la technique suivante : (i) produire autant de petits papiers identiques qu'il y a d'élèves inscrits dans la classe, (ii) sur 28 d'entre eux, noter un symbole, les autres papiers restant vides, (iii) mélanger les papiers face cachée et noter au dos de chacun d'eux le nom d'un élève de la classe. Les élèves dont le papier est marqué du symbole sont sélectionnés. Le nombre d'élèves qui seront effectivement évalués peut donc être égal ou inférieur à 28 selon le hasard de l'attribution. Les papiers marqués du symbole et attribués à des élèves absents seront comptabilisés comme des non-réponses. Il convient, comme précédemment, de documenter le nombre total d'absent et le nombre d'absents qui auraient dû être testés.

5 Il s'agit de 30 élèves en fin de collège (9^e/10^e année).

3.2. Documentation

Comme il vient d'être mentionné, sans informations précises relatives à l'échantillonnage de la classe et des élèves, la pondération finale des élèves ne pourra être calculée. L'administrateur de test doit donc soigneusement consigner (i) le nombre de classes du niveau cible dans l'école, (ii) le nombre d'élèves inscrits dans la classe sélectionnée (en indiquant le nombre de présents et d'absents), (iii) le nombre d'élèves réellement testés et (iv) le nombre de non-réponses, c'est-à-dire les élèves sélectionnés mais absents ce jour. Elles seront consignées dans le *Formulaire de suivi de l'école* (en anglais, *School Tracking Form*) et dans le *Formulaire de suivi des élèves* (en annexe). Ces formulaires sont d'une grande importance.

3.2.1. Le formulaire de suivi de l'école

Le *Formulaire de suivi de l'école* est un document unique à chaque école et à chaque échantillon. À une école sélectionnée pour les deux niveaux correspond deux formulaires de suivi distincts : un pour la sixième primaire et un pour la deuxième. Nous conseillons que le centre national édite chaque *Formulaire de suivi de l'école* avant l'évaluation et y consigne les informations suivantes :

- l'identifiant national de l'école ;
- l'identifiant PASEC de l'école ;
- le nom de l'école et l'adresse ;
- une indication du nombre d'effectifs pour le grade visé (SIZE6 ou SIZEGRAD) ;

Sur place, le jour de l'évaluation, l'administrateur de test complète les données suivantes :

- le code identifiant de l'administrateur de test : il est important de pouvoir garder trace de l'administrateur de test pour chaque école évaluée. Chaque administrateur de test recevra donc un identifiant unique, un code à deux chiffres par exemple ;
- la date de la séance ;
- le nombre total d'élèves inscrits en 6^e année du primaire (y compris les absents depuis moins d'un mois) ;
- le nombre de classes organisées pour la 6^e année du primaire. Si l'école fonctionne en double vacation, la classe du matin et la classe de l'après-midi doivent être considérées comme deux classes distinctes ;
- le nombre d'élèves inscrits (y compris les absents depuis moins d'un mois) dans la classe sélectionnée ;
- le nombre total d'élèves exclus ;
- les nombres d'élèves présents et absents à chacune des sessions du test ;
- la réception du questionnaire enseignant complété ;
- la réception du questionnaire directeur complété.

3.2.2. Le formulaire de suivi des élèves

Le *Formulaire de suivi des élèves* sert à enregistrer des données concernant les élèves échantillonnés, y compris les élèves absents un ou plusieurs jours d'évaluation. Sont consignées, des informations générales relatives à l'évaluation et des informations pour chaque élève de l'échantillon.

Les informations générales sont :

- la date du premier jour d'évaluation ;
- l'identifiant PASEC de l'école ;
- le nom de l'école ;
- l'identifiant de l'administrateur de test ;

Les informations relatives à chaque élève sont :

- un identifiant élève pré-imprimé : 28 ou 30 identifiants sont prévus (16 pour la 2^e) respectivement pour la 6^e année du primaire et 9^e/10^e année (fin du collège). Si l'école comprend moins de 28/30 élèves du niveau, seules les lignes correspondant au nombre réel d'élèves échantillonnés seront complétées ;
- le niveau scolaire ;
- les nom et prénom de l'élève ;
- le sexe de l'élève ;
- la date de naissance de l'élève si cette information est disponible ;
- l'éventuelle exclusion et sa justification ;
- la participation aux différentes sessions de l'évaluation ;
- la forme du carnet, pré-imprimée.

Si les deux grades 6^e année et 2^e année sont concernés par l'étude, deux formulaires distincts sont complétés. Les modèles des *Formulaires de suivi des élèves* 6^e année et 2^e année du primaire sont proposés en annexe.



1

2

3

4

5

Chapitre 4

Pondération des données



La pondération est d'abord décrite pour la sixième année primaire. En fin de section seront décrites les modifications à apporter à la procédure pour la pondération des données de deuxième année (Martin et al., 2003).

4.1. Préparation des fichiers

Avant de pondérer des données, il importe de procéder à un ensemble de vérifications des différents fichiers au départ desquels ce travail sera réalisé.

À ce stade, le centre de coordination international devrait disposer, par pays, de deux fichiers STATA, élaborés lors du tirage des écoles, à savoir :

1. la *Base de sondage* qui reprend l'ensemble des écoles du système éducatif, après suppression des écoles exclues ;
2. par année d'études, le fichier *Échantillon* qui reprend les écoles échantillonnées et les premières et deuxièmes écoles de remplacement. Dans ce fichier figurent, outre les identifiants nationaux et PASEC, les variables de stratification (implicites et explicites), les variables de taille et le statut (école échantillonnée 'S', école de remplacement 1 'R1' ou école de remplacement 2 'R2') et la probabilité de sélection pour les deux années d'études.

Le second fichier a déjà fait l'objet de nombreuses vérifications avant d'être envoyé au centre national, avant la campagne d'évaluation. Après la collecte des données sur le terrain, le codage et la saisie des données sur feuilles optiques, le centre national transmet au centre de coordination internationale, et ce pour chaque année d'études, l'ensemble des données.

Parmi ces données, deux fichiers sont d'une extrême importance pour le processus de pondération. Ces deux fichiers sont :

1. le fichier électronique contenant les *Formulaires de suivi des élèves* pour l'ensemble des écoles participantes ;
2. le fichier *Liste des écoles* qui reprend, pour l'ensemble des écoles participantes et les écoles échantillonnées qui n'ont pas participé, qu'elles aient ou non été remplacées, quelques informations qui seront énumérées ultérieurement. Ce fichier a été construit sur base des informations reprises dans les *Formulaires de suivi des écoles*.

Avant de confronter ces deux nouveaux fichiers de données avec le fichier *Échantillon*, il importe de vérifier l'intégrité et la consistance de ces fichiers avec les fichiers *Test de Français Élèves* et *Test de Mathématiques Élèves*, ainsi que les données des fichiers *Questionnaire élèves*.

Cette vérification doit être implémentée séparément pour le test de français, de mathématiques et pour le questionnaire contextuel puisqu'il est recommandé que la lecture optique soit organisée pour saisir ces données dans trois fichiers distincts.

Pour tout élève rapporté présent pour l'une des deux sessions d'un test cognitif ou pour la session du questionnaire, un enregistrement avec le même identifiant devrait être trouvé dans le fichier de données correspondant. En cas de discordance, il importe de consulter le centre national pour qu'il opère une vérification et qu'il solutionne le problème. Par ailleurs, un élève qui serait considéré comme exclu mais présent avec des données constitue un autre type d'incohérence. Toutes ces vérifications seront largement décrites dans la section VI de ce document, exclusivement consacrée au nettoyage de données et reprise dans le manuel des centres nationaux.

À ce stade, nous considérons que le nettoyage des données est clôturé. En d'autres termes, le fichier *Formulaires de suivi des élèves* et les trois fichiers de *Données élèves* sont consistants, c'est-à-dire que :

1. tout élève noté présent lors d'une des deux sessions d'un test cognitif ou lors de la session consacrée au questionnaire est repris dans le fichier de données correspondant à l'instrument. Si l'élève était présent lors de la première session et absent lors de la seconde session d'un test cognitif ou inversement, des codes *missing* ont été saisis pour la session au cours de laquelle il a été rapporté absent ;
2. tout élève exclu est rapporté comme tel dans le *Formulaire de suivi des élèves* et aucun enregistrement relatif à cet élève n'est trouvé dans les fichiers de *Données Élèves* ;
3. tout élève rapporté absent pour l'un des deux tests ou le questionnaire n'est pas enregistré dans le fichier de données correspondant.

TABLEAU 20 : Profils possibles des quatre variables après fusion des fichiers *Échantillon* et *Liste des écoles*

STATUT6	PARTICIPE6	ECHAN6	LDE6	
S	1	1	1	École échantillonnée participante
S	0	1	1	École échantillonnée non participante
R1	1	1	1	École de remplacement R1 participante
R1	0	1	0	École de remplacement R1 non participante
R2	1	1	1	École de remplacement R2 participante
R2	0	1	0	École de remplacement R2 non participante

Tout autre pattern n'est pas compatible avec la méthodologie proposée. En effet :

1. les écoles de remplacement qui n'ont pas participé ne peuvent pas figurer dans le fichier *Liste des écoles*, peu importe que ces écoles n'aient pas été sollicitées ou qu'elles aient refusé en cas de sollicitation ;
2. la variable ECHAN6 ne peut avoir la valeur 0. Dans le cas contraire, un identifiant école est erroné dans le fichier *Liste des écoles* ;
3. pour les écoles échantillonnées (STATUT6=S), les variables ECHAN6 et LDE6 ne peuvent prendre que la valeur 1 ;
4. pour les écoles de remplacement R1 et R2, si PARTICIPE6=1, alors LDE6 doit être égal à 1 et si PARTICIPE6=0, alors LDE6 doit être égal à 0.

Si le fichier *Liste des écoles* issu de cette fusion répond aux normes susmentionnées, toutes les écoles de remplacement qui n'ont pas participé à l'enquête peuvent être supprimées. À ce stade, il reste dans le fichier *Liste des écoles*, les écoles échantillonnées et les écoles de remplacement qui ont participé.

Le deuxième travail consiste à vérifier qu'une école échantillonnée et une de ses écoles de remplacement n'aient pas participé toutes les deux. Comme les écoles de remplacement ont reçu un identifiant qui permet de déterminer l'identifiant de l'école échantillonnée correspondante, (pour rappel, plus 300 pour R1 et plus 600 pour R2), on peut aisément identifier une double participation, en structurant les données comme le *Tableau 21*. Ce tableau doit être réalisé séparément pour les écoles de l'échantillon qui n'ont pas été exclues après échantillonnage et les écoles de l'échantillon qui ont été exclues.

TABLEAU 21 : Écoles échantillonnées non exclues et écoles de remplacement

Original – école S			Remplacement 1 – école R1			Remplacement 2 – école R2		
ID_STRATE	ID_PASEC	PARTICIPE6	ID_STRATE	ID_PASEC	PARTICIPE6	ID_STRATE	ID_PASEC	PARTICIPE6
01	011001	1						
01	012002	0	01	012302	1			
01	012003	0				01	012603	1
01	011004	1	01	011004	1			

Si l'école échantillonnée a participé, il n'y a aucune raison que les données relatives à ses deux écoles de remplacement apparaissent. En présence d'une telle erreur, il importe de vérifier que le code de participation de l'école échantillonnée est correct, auquel cas l'école de remplacement sera supprimée du fichier. Le deuxième exemple du *Tableau 21* présente le cas d'une école échantillonnée qui refuse de participer et dont la première école de remplacement accepte de participer. Le troisième exemple est identique au second, à l'exception que la première école de remplacement a également refusé de participer ; par contre, la seconde école de remplacement a accepté. Il convient de noter dans ce cas que la première école de remplacement n'apparaît pas dans le fichier. Imaginons que la deuxième école de remplacement ait aussi refusé de participer. Dans ce cas, seules les données de l'école échantillonnée seraient reprises dans cet enregistrement. Enfin, le quatrième exemple illustre une erreur puisque l'école échantillonnée et la première école de remplacement ont toutes les deux participé. Comme indiqué précédemment, il importe de vérifier auprès des centres nationaux cette incohérence avant de (i) supprimer l'école de remplacement ou (ii) de corriger le code de participation de l'école échantillonnée. Le responsable de la pondération consultera la version électronique du fichier *Formulaires de suivi des élèves* pour déterminer si des données élèves existent pour les deux écoles ou pour une seule des deux écoles.

Pour le tableau qui reprend uniquement les écoles exclues, seules des écoles échantillonnées devraient apparaître puisqu'aucun remplacement ne peut être effectué dans ce cas.

Si aucune erreur ne subsiste dans ce fichier *Liste des écoles*, alors le nombre total d'écoles (écoles échantillonnées 'S', indépendamment de leur participation, plus les écoles de remplacement 'R1', 'R2' qui ont participé) ne peut dépasser le nombre total d'écoles de l'échantillon initial plus le nombre d'écoles échantillonnées qui n'ont pas participé, hormis les exclusions. Imaginons un échantillon national de 180 écoles. Il s'avère, lors des opérations de terrains, que trois écoles auraient dû être exclues. En effet, la première est une école privée qui ne dépend pas du ministère de l'éducation, elle n'appartient donc pas à la population cible. La deuxième n'est fréquentée que par des élèves relevant d'une des catégories d'exclusion. Enfin, la troisième école a définitivement fermé, faute d'élèves. Ces trois écoles ne peuvent bien entendu être remplacées. L'échantillon ne compte donc plus théoriquement que 197 écoles potentiellement participantes. Imaginons encore que 5 écoles de cet échantillon refusent de participer. Ces 5 dernières pourront quant à elles faire l'objet d'un remplacement. Le fichier *Liste des écoles*, à ce stade des opérations, ne peut donc pas dépasser les 205 enregistrements.

Dans la troisième phase du travail, il convient ensuite de vérifier si les informations contenues dans le fichier *Liste des écoles* sont bien cohérentes avec les informations contenues dans le fichier *Formulaires de suivi des élèves*. Étant donné que la cohérence entre le fichier *Formulaires de suivi des élèves* et les fichiers de *Données Tests* et *Données Questionnaire* a été confirmée lors du nettoyage des données, ces vérifications porteront notamment mais pas de manière exhaustive sur les points listés ci-dessous.

1. Pour chaque école participante, le nombre d'enregistrements dans le fichier *Formulaires de suivi des élèves* doit être égal à NB_SAMP6 (le nombre d'élèves échantillonnés).
2. Au départ du fichier *Formulaires de suivi des élèves*, calculer, par école, le nombre d'élèves présents pour au moins une session, le nombre d'élèves exclus et le nombre d'élèves absents (c'est-à-dire absents à toutes les sessions). Ces nombres devraient être identiques aux valeurs reprises dans le fichier *Liste des écoles* (pour rappel, respectivement NB_PRE6, NB_EX6 et NB_AB6). En cas de discordance, il importe de consulter le centre national.
3. Une école échantillonnée non participante ou exclue ne peut avoir d'enregistrements dans le fichier *Formulaires de suivi des élèves*.
4. Une école rapportée comme participante avec aucun enregistrement dans le fichier *Formulaires de suivi des élèves* constitue aussi une incohérence. Le centre national sera consulté pour résoudre ce problème.
5. En cas de cohérence entre les nombres d'élèves participants, d'élèves absents et d'élèves exclus tels que rapportés dans le fichier *Liste des écoles* et le fichier *Formulaires de suivi des élèves*, si 25 ou moins de 25 pourcent des élèves ont participé (soit généralement moins de 5), l'école sera considérée comme non participante et les données de ces élèves seront supprimées des différents fichiers (*Test*, *Questionnaire élèves*, *Questionnaire maîtres*, *Questionnaire écoles*, *Formulaires de suivi des élèves* et *Liste des écoles*).

Après ce précieux travail de vérifications de cohérences multiples et de nettoyage, le fichier *Liste des écoles* est prêt pour implémenter la pondération.

4.2. Calcul du poids initial de l'école

La pondération initiale de l'école (W_SCH_INIT) est égale à (Martin et al., 2003) :

$$W_SCH_INIT_i = \frac{1}{PROB_i}$$

Par ailleurs, une seconde variable de type indicatrice (dichotomique 0/1) sera également créée pour différencier les écoles certaines des écoles non certaines. Cette différenciation ne devrait poser aucune difficulté puisque les écoles certaines ont été rassemblées dans une strate explicite spécifique.

Ces deux opérations réalisées, plusieurs nouvelles vérifications doivent être implémentées, à savoir :

1. les écoles certaines ont toutes une probabilité de sélection et un poids initial égaux à 1 ;
2. les écoles non certaines ont une probabilité de sélection inférieure à 1 mais supérieure à 0 et un poids supérieur à 1 ;
3. au sein d'une strate explicite, le produit $W_SCH_INIT6 * MOS$ pour les seules écoles échantillonnées correspond à la somme des MOS pour cette même strate explicite ;
4. le nombre d'élèves de 6^e année dans la *Base de sondage* ne diffère pas de plus de 15 pourcent du nombre d'élèves de 6^e tel qu'il apparaît dans le fichier *Liste des écoles*. Dans le cas contraire, il sera demandé au centre national d'opérer les vérifications qui s'imposent ;

Pour les seules écoles échantillonnées (écoles 'S'), on calcule par école :

- le produit $W_SCH_INIT6 * TOT_ELS6$ et la somme de ces produits par strate explicite ;
- le produit équivalent mais avec le nombre d'élèves tel que mentionné dans la *Base de sondage* ($SIZE6$), soit $W_SCH_INIT6 * SIZE6$ et la somme de ces produits par strate explicite ;

Toute différence entre les deux sommes devrait en principe traduire une variation dans la taille de la population spécifique à cette strate. Il importe d'interroger le centre national si une différence de plus de 10 pourcent existe entre ces deux sommes.

4.3. Calcul du taux de participation des écoles

À ce stade, le fichier école comporte :

- toutes les écoles échantillonnées. Il s'agit à la fois des écoles qui ont participé, de celles qui ont refusé et enfin de celles qui ont été exclues après échantillonnage ;
- les écoles de remplacement qui ont participé.

Tout comme les études de l'IEA ou l'étude PISA, il importe de calculer des taux de participation des écoles, pondérés et non pondérés. De plus, il est d'usage de distinguer les taux de réponse (i) avant prise en considération des écoles de remplacement, (ii) avec les écoles R1 et enfin (iii) avec les écoles R1 et R2. En effet, bien que les règles varient d'un organisme à l'autre, voire même d'une étude à l'autre au sein d'un organisme, l'analyse des taux de réponses écoles, élèves et/ou des taux globaux conduit généralement à classer l'échantillon national dans une des trois catégories suivantes :

1. l'échantillon est adjudiqué sans aucune réserve ;
2. l'échantillon est adjudiqué avec quelques réserves ; on invite de la sorte le lecteur à interpréter prudemment les résultats ;
3. l'échantillon n'est pas adjudiqué.

L'usage de cette adjudication varie également selon les organismes. Les données adjudiquées dans PISA, avec ou sans réserve, sont publiées et rendues publiques sans aucune distinction. Les données non adjudiquées ne sont par contre pas publiées. Les responsables de l'étude PISA demande aux centres nationaux dont l'échantillon a été classé dans la seconde catégorie de procéder à des analyses statistiques pour identifier d'éventuels biais dans l'échantillon. Il s'agit essentiellement de procéder à des régressions logistiques avec comme variable dépendante, la participation *versus* non-participation de l'école, et comme variables indépendantes, le niveau socio-économique de l'école, le caractère urbain ou non de l'école, la distinction public-privé... Les variables indépendantes modélisées dépendent bien entendu du pays. Si aucun biais ne peut être identifié, les données sont publiées sans restriction. Dans le cas contraire, les données ne seront pas publiées.

Dans le cadre des études de l'IEA, toutes les données sont publiées, mais soit avec une notice pour les données adjudiquées avec réserves, soit en annexes ou en bas de tableaux pour les données non adjudiquées.

Dans le cadre de l'étude PISA, le taux de participation des écoles exigé avant remplacement s'élève à 85 %, et le taux de participation exigé après remplacement, pour autant que le taux initial soit supérieur à 65 % pourcent, s'élève à $85 + (85 - \text{taux avant remplacement})/2$. Ainsi, si le taux de réponse des écoles échantillonnées s'élève à 75 %, alors le pays devra atteindre, selon PISA, un taux de participation de $85 + (85 - 75)/2 = 90$ %.

À titre d'exemple, les études TIMSS et PIRLS exigent actuellement un taux pondéré ou non pondéré de 85 % des écoles avant remplacement et un taux de 85 % des élèves pour être repris dans la première catégorie.

Quelles que soient les règles d'adjudication adoptées, les taux de réponse Écoles et Élèves doivent être publiés dans les rapports internationaux.

Pour calculer les taux de participation, il importe préalablement de mettre de côté toutes les écoles qui ont été exclues, après bien entendu avoir vérifié correctement qu'aucune école de remplacement ne soit associée à ces écoles exclues.

Par ailleurs, pour calculer le taux pondéré, il est nécessaire de disposer du poids initial de l'école, c'est-à-dire W_{SCH_INIT6} et d'une mesure de taille. Il est d'usage d'utiliser le nombre d'élèves de l'année d'étude tel que rapporté dans la *Base de sondage* ou dans le fichier *Échantillon* (SIZE6). Si cette information n'est pas disponible, alors on privilégiera le nombre d'élèves de 6^e année repris dans le fichier *Liste des écoles* (TOT_ELS6). Enfin, si aucune de ces deux informations n'est disponible, on utilisera la taille moyenne par grade (SIZEGRAD) issue de la *Base de sondage*.

TABLEAU 22 : Écoles de l'échantillon initial et écoles de remplacement participantes

École	Statut	SIZE6	MOS	Prob	Poids	Poids × MOS	Poids × SIZE6	Exclusion	Participation
1	S	30	25	0,05	20,00	500,00	600,0	0	1
2	S	32	30	0,06	16,67	500,00	533,3	0	1
3	S	27	35	0,07	14,29	500,00	385,7	0	0
303	R1	39	38	0,08	13,16	500,00	513,2	0	1
4	S	46	40	0,08	12,50	500,00	575,0	0	0
304	R1	37	42	0,08	11,90	500,00	440,5	0	1
5	S	41	45	0,09	11,11	500,00	455,6	0	1
6	S	53	50	0,10	10,00	500,00	530,0	0	0
606	R2	58	53	0,11	9,43	500,00	547,2	0	1
7	S	50	55	0,11	9,09	500,00	454,5	0	1
8	S	62	60	0,12	8,33	500,00	516,7	0	1
9	S	68	65	0,13	7,69	500,00	523,1	0	1
10	S	55	70	0,14	7,14	500,00	392,9	0	0
310		67	72	0,14	6,94	500,00	465,3	0	1
11	S	80	75	0,15	6,67	500,00	533,3	0	1
12	S	82	80	0,16	6,25	500,00	512,5	0	0
13	S	82	85	0,17	5,88	500,00	482,4	0	1
14	S	95	90	0,18	5,56	500,00	527,8	0	0
15	S	90	95	0,19	5,26	500,00	473,7	0	1
16	S	100	100	0,20	5,00	500,00	500,0	0	1
17	S	100	105	0,21	4,76	500,00	476,2	0	1
18	S	115	110	0,22	4,55	500,00	522,7	0	1
19	S	113	115	0,23	4,35	500,00	491,3	1	0
20	S	127	120	0,24	4,17	500,00	529,2	1	0

Les calculs présentés ci-dessous seront illustrés au départ d'un exemple basé sur le *Tableau 22*. L'échantillon initial de 20 écoles a été tiré d'une population dont la somme des mesures de taille MOS s'élève à 10 000. Pour rappel, dans le cadre du PASEC, la mesure de taille MOS d'une école peut être généralement assimilée à la moyenne du nombre d'élèves en deuxième et en sixième. Les 20 écoles ont été sélectionnées avec une probabilité proportionnelle à cette mesure de taille. Toutefois, le fichier de données comporte 24 écoles. En effet, six écoles ont refusé de participer ou n'ont pas pu participer. Trois écoles (3, 4 et 10) ont été remplacées par leur première école de remplacement, une (école 6) par sa deuxième école de remplacement et deux écoles (écoles 12 et 19) n'ont pas été remplacées.

Par ailleurs, les écoles 19 et 20 ont été exclues après échantillonnage, peu importe les raisons de cette exclusion. Ces deux écoles peuvent être mises sur le côté temporairement pour le calcul des taux de participation. Par contre, ces écoles seront intégrées dans le calcul des taux d'exclusion qui doivent également être rapportés dans les annexes techniques des rapports internationaux.

Les calculs présentés ci-dessous se basent donc sur 18 écoles de l'échantillon initial, 3 écoles de remplacement 1 et une école de remplacement 2.

Le *Tableau 22* fournit également la probabilité de sélection et le poids initial.

Comme les écoles ont été tirées avec une probabilité proportionnelle à leur taille, le produit poids initial * MOS est une constante. En effet,

$$\frac{1}{\frac{MOS_i n_{sch}}{MOS}} MOS_i = \frac{MOS}{MOS_i n_{sc}} MOS_i = \frac{MOS}{n_{sc}} = \frac{10000}{20} = 500$$

Par contre, comme cette mesure de taille synthétise les effectifs de deuxième et de sixième, le produit poids initial (W_{SCH_INIT6}) avec la taille des effectifs de sixième ($SIZE6$) n'est plus une constante.

4.3.1. Taux de participation des écoles avant remplacement

Pour calculer le taux de participation des écoles avant remplacement, toutes les écoles de remplacement peuvent être supprimées du fichier. Ainsi, dans l'exemple ci-dessus, les écoles 303, 304, 606 et 310 seront éliminées.

Le taux de participation non pondéré des écoles avant remplacement est égal au nombre d'écoles de l'échantillon initial qui ont participé, divisé par le nombre d'école de l'échantillon initial qui ont ou auraient dû participer, le tout multiplié par 100.

Dans l'exemple, ce taux sera donc égal à :

$$T_R_NP_EC_I = 100 \left[\frac{12}{18} \right] = 66,6\%$$

Le taux de participation pondéré des écoles avant remplacement implique les mêmes enregistrements et prend en considération le poids de l'école et les effectifs de sixième repris de la *Base de sondage*. Si le taux de participation pondéré était basé sur la mesure de taille (MOS), alors les taux de participation pondéré et non pondéré seraient identiques puisque les produits $W_{SCH_INIT6} * MOS$ sont tous égaux. Par contre, l'usage de $SIZE6$ en lieu et place de MOS génère une certaine variabilité dans les produits qui peut générer une différence entre taux pondéré et taux non pondéré.

Le taux de participation pondéré des écoles avant remplacement est donc égal à la somme des produits $W_{SCH_INIT6} * SIZE6$ pour les écoles de l'échantillon initial qui n'ont pas été exclues et qui ont participé, divisé par la somme de ces mêmes produits pour l'ensemble des écoles de l'échantillon initial qui n'ont pas fait l'objet d'une exclusion, le tout multiplié par 100.

$$\text{Dans l'exemple, } T_R_P_EC_I = 100 \left[\frac{6071,5}{8995,3} \right] = 67,5\%.$$

Comme on peut le constater, dans l'exemple, taux pondéré et non pondéré sont relativement proches.

4.3.2. Taux de participation des écoles avec les premières écoles de remplacement

Le calcul des taux de participation des écoles avec les premières écoles de remplacement suppose qu'au niveau du numérateur, les premières écoles de remplacement soient comptabilisées au même titre que les écoles de l'échantillon initial non exclues qui ont participé et qu'au dénominateur, les premières écoles de remplacement soient aussi comptabilisées, mais en lieu et place des écoles qu'elles remplacent. En d'autres termes, les écoles de l'échantillon initial qui ont été remplacées par leur première école de remplacement n'interviennent plus au dénominateur.

Dans l'exemple, les écoles 303, 304 et 310 contribueront au calcul du numérateur et remplaceront les écoles 3, 4 et 10 au dénominateur.

Dans l'exemple :

$$T_R_NP_EC_{R1} = 100 \left[\frac{15}{18} \right] = 83,3\%$$

$$T_R_P_EC_{R1} = 100 \left[\frac{7490,4}{9060,7} \right] = 82,7\%$$

4.3.3. Taux de participation des écoles avec les premières et deuxièmes écoles de remplacement

Les calculs opérés pour les taux avec les premières écoles de remplacement intègrent à présent les deuxièmes écoles de remplacement.

$$T_R_NP_EC_{R2} = 100 \left[\frac{16}{18} \right] = 88,8\%$$

$$T_R_P_EC_{R2} = 100 \left[\frac{8037,5}{9077,8} \right] = 88,6\%$$

Dans l'exemple repris ci-dessus, les taux de participation pondéré et non pondéré sont très semblables. Ce cas de figure est relativement classique avec un échantillon pour lequel les écoles sont tirées avec une probabilité proportionnelle à la taille de l'établissement. Des différences substantielles entre ces deux taux devraient susciter une investigation approfondie pour en connaître l'origine car ces différences pourraient notamment témoigner d'une contribution excessive d'une ou de plusieurs écoles.

4.3.4. Taux d'exclusion des écoles de l'échantillon initial

Il importe aussi de calculer ou d'estimer les différents types d'exclusion, en fonction du moment et du lieu de ces exclusions.

Pour rappel, il est possible, lors de la construction de la *Base de sondage*, d'exclure des écoles à deux moments, pour diverses raisons :

1. une réduction de la couverture de la population, essentiellement pour des raisons politiques ou linguistiques ;
2. l'exclusion, au sein de la population couverte, des écoles qui n'accueillent que des élèves qui relèvent d'une des catégories d'exclusion d'élèves ;
3. l'exclusion des petites écoles, pour autant que ces exclusions ne représentent pas plus d'1% de la population d'élèves.

On peut à présent estimer le pourcentage d'écoles exclues lors des opérations de terrain.

Le taux non pondéré d'écoles exclues lors des opérations de terrain s'élève donc au nombre d'écoles exclues, divisés par le nombre d'écoles de l'échantillon initial, soit dans notre exemple $\frac{2}{20} \cdot 100 = 10\%$. Le taux pondéré doit donc faire intervenir,

tant au numérateur qu'au dénominateur les sommes des produits $W_SCH_INIT6 * SIZE6$. Au numérateur, seules les écoles exclues interviennent. Au dénominateur, toutes les écoles de l'échantillon initial seront comptabilisées.

Dans notre exemple, ce taux d'exclusion pondéré sera donc égale à $\frac{1020,5}{10015,8} * 100 = 10,2\%$.

4.4. Calcul de l'ajustement pondéral pour la non-réponse des écoles

Dans la plupart des pays participants, et comme l'illustre l'exemple du *Tableau 22*, des écoles peuvent refuser de participer à l'enquête. Dans certains pays occidentaux, les taux de participation des écoles sont tellement faibles qu'ils peuvent conduire à la non-publication des résultats et des données. Dans le contexte africain, et comme en témoigne certains des rapports d'évaluation des études PASEC menées dans les années 90 et 2000, ces refus constituent plutôt un phénomène marginal.

Quoi qu'il en soit, et indépendamment de son importance, la non-réponse est susceptible de générer un biais dans les résultats (Ardilly, 2006). La meilleure façon pour minimiser les biais est, bien entendu, de déployer, au niveau des centres nationaux, tous les moyens pour obtenir le taux de participation le plus élevé. Il serait à cet égard erroné de se contenter d'expliquer un faible taux de participation par une culture nationale de non-participation ou de refus et ainsi minimiser le rôle des autorités et des coordinateurs nationaux. Ainsi, le taux de participation des écoles au Pays-Bas avoisinait les 27 % dans PISA 2000 alors qu'il dépassait les 80 % en 2003.

Au-delà des efforts consentis par les centres nationaux, les enquêtes internationales à large échelle en éducation ont pour habitude d'implémenter deux procédures conçues pour réduire les biais potentiels engendrés par la non-réponse des écoles :

1. à chaque école de l'échantillon initial, sauf à quelques exceptions près et plus particulièrement dans les pays de petites tailles, sont associées deux écoles de remplacement ;
2. un ajustement pondéral.

Si toutes les écoles qui ne peuvent ou qui refusent de participer sont suppléées par leur première ou leur deuxième école de remplacement, le taux de participation avec les écoles de remplacement s'élèverait à 100 % et dès lors, il n'y aurait pas lieu d'opérer un ajustement pondéral pour la non-réponse. Cet ajustement est donc uniquement conçu pour compenser la non-participation des écoles de l'échantillon initial qui n'ont pas pu être remplacées. Dans l'exemple du *Tableau 22*, il s'agit des écoles 12 et 14. Aucun ajustement ne sera appliqué pour les écoles 3, 4, 6 et 10 puisque ces dernières ont été remplacées.

Les différences entre les études classiques de l'IEA, à savoir PIRLS et TIMSS et l'étude PISA de l'OCDE ont déjà été évoquées à plusieurs reprises, parmi lesquelles on peut rappeler :

1. population d'âge versus population définie en termes d'une année d'études ;
2. échantillon aléatoire et simple d'une ou de plus d'une classe de l'année d'études investiguée *versus* échantillon aléatoire et simple de 35 élèves de 15 ans au sein des écoles, toute classe et toute année d'études confondues ;
3. estimation de la variance d'échantillonnage par la variante FAY *versus* par la méthode JK2 des *Balanced Repeated Replication*.

Une nouvelle distinction s'opère également à ce niveau. En effet, dans les études de l'IEA, comme il sera évoqué ultérieurement dans la section relative à l'estimation de la variance par des méthodes de ré-échantillonnage, il est nécessaire de constituer des paires d'écoles au sein des strates explicites. Dans le cadre des études PIRLS et TIMSS, ces paires sont constituées uniquement au départ des écoles participantes. La procédure peut être décrite comme suit : l'ajustement pondéral pour la non-réponse des écoles est implémenté, ces écoles sont supprimées du fichier et les paires d'écoles sont constituées au départ uniquement des écoles participantes. Il s'ensuit que l'estimation de la variance d'échantillonnage dans les études TIMSS et PIRLS ne prend pas en compte la non-réponse des écoles.

L'étude PISA procède différemment. En effet, un ajustement pondéral pour la non-réponse des écoles est calculé pour chacune des réplifications. La variance d'échantillonnage dans PISA intègre donc une incertitude supplémentaire.

Dans le cadre de cette section, seule l'approche adoptée par l'IEA, qui sera également la procédure PASEC, sera décrite. Toutefois, comme on vient de le mentionner, cette approche ne peut être généralisée à l'étude PISA, même si conceptuellement, les principes fondamentaux de l'ajustement pondéral restent identiques. Plus précisément, il s'agit de répartir ce qu'aurait dû être la contribution des écoles non participantes sur les écoles participantes. Autrement dit, le facteur d'ajustement est égal au rapport entre d'une part la contribution des écoles participantes et des écoles qui auraient dû participer, et d'autre part, les écoles participantes.

Repartons du *Tableau 22* pour illustrer cet ajustement pondéral pour la non-réponse des écoles. Son calcul nécessite la suppression dans la base de données :

1. des écoles qui ont été exclues, puisque ces écoles ne figurent plus dans la population que l'échantillon est censé représenter (les écoles 19 et 20 dans l'exemple) ;
2. des écoles de l'échantillon initial qui, soit ont refusé de participer, soit, n'ont pas pu participer ET qui ont été remplacées (les écoles 3, 4, 6 et 10).

Le *Tableau 23* reprend les écoles au départ desquelles l'ajustement pour la non-réponse des écoles sera calculé. Il reprend les écoles participantes et les écoles de l'échantillon initial non participantes et qui n'ont pas été remplacées.

TABLEAU 23 : Écoles participantes et les écoles de l'échantillon initial non participantes et qui n'ont pas été remplacées

École	Statut	SIZE6	MOS	Prob	Poids	Poids × SIZE6	Exclusion	Participation	Poids ajusté	Poids ajusté × SIZE6
1	S	30	25	0,05	20,00	600,0	0	1	22,6	677,7
2	S	32	30	0,06	16,67	533,3	0	1	18,8	602,4
303	R1	39	38	0,08	13,16	513,2	0	1	14,9	579,6
304	R1	37	42	0,08	11,90	440,5	0	1	13,4	497,5
5	S	41	45	0,09	11,11	455,6	0	1	12,5	514,5
606	R2	58	53	0,11	9,43	547,2	0	1	10,7	618,0
7	S	50	55	0,11	9,09	454,5	0	1	10,3	513,4
8	S	62	60	0,12	8,33	516,7	0	1	9,4	583,5
9	S	68	65	0,13	7,69	523,1	0	1	8,7	590,8
310	R1	67	72	0,14	6,94	465,3	0	1	7,8	525,5
11	S	80	75	0,15	6,67	533,3	0	1	7,5	602,4
12	S	82	80	0,16	6,25	512,5	0	0	7,0	574
13	S	82	85	0,17	5,88	482,4	0	1	6,6	544,8
14	S	95	90	0,18	5,56	527,8	0	0	6,2	589
15	S	90	95	0,19	5,26	473,7	0	1	5,9	535,0
16	S	100	100	0,20	5,00	500,0	0	1	5,6	564,7
17	S	100	105	0,21	4,76	476,2	0	1	5,4	537,8
18	S	115	110	0,22	4,55	522,7	0	1	5,1	590,4

Ce fichier comporte 18 enregistrements, ce qui correspond à la taille de l'échantillon initial moins les écoles exclues.

L'ajustement pondéral pour la non-réponse des écoles prend en considération le poids initial de l'école et la taille de l'école. Comme dans le cadre du calcul du taux de participation, il est conseillé d'utiliser le nombre d'élèves de l'année d'étude tel que rapporté dans la *Base de sondage* ou dans le fichier *Échantillon* (SIZE6). Si cette information n'est pas disponible, alors on privilégiera le nombre d'élèves de 6^e année repris dans le fichier *Liste des écoles* (TOT_ELS6). Enfin, si aucune de ces deux informations n'est disponible, on prendra la mesure de taille moyenne (SIZEGRAD).

La première étape de l'ajustement pondéral pour la non-réponse des écoles, après bien entendu avoir supprimé les écoles exclues et les écoles de l'échantillon initial qui ont été remplacées, consiste à définir des groupes d'écoles au sein desquels cet ajustement sera implémenté. En effet, si aucun groupe n'est défini, le même ajustement pondéral sera appliqué à l'ensemble des écoles participantes. Si l'on multiplie le poids de toutes les écoles participantes par un coefficient de 1.25 par exemple, alors les estimations avant et après ajustement seront parfaitement identiques, à l'exception de la somme des pondérations (et de tout autre indice dérivé de cette somme comme la moyenne des poids).

Pour que l'ajustement pondéral puisse réduire le biais introduit par la non-réponse, il importe de définir les groupes d'écoles (i) en fonction de variables qui corréleront avec la performance et (ii) pour lesquels des différences de participation sont observées. En effet, si dans chaque groupe, le taux de participation est de 80 %, l'ajustement sera identique pour tous les groupes et dans ce cas, comme on vient de le voir, il n'affectera pas les estimations.

Les études TIMSS et PIRLS utilisent les strates explicites définies lors de la conception du plan d'échantillonnage comme groupes pour le calcul de la non-réponse des écoles. Ce choix présente l'avantage qu'aucun travail supplémentaire n'est demandé et, par ailleurs, s'avère généralement pertinent puisque les variables de stratification ont été choisies en partie en fonction de leur corrélation avec la performance.

Dans le cadre de PISA, ces groupes sont constitués au sein des strates explicites en fonction des variables de stratification implicite. En théorie, cette méthodologie permet de mieux contrebalancer le biais potentiel de la non-réponse mais il nécessite un surcroît de travail important. Pour cette raison, il est conseillé au PASEC d'adopter la méthodologie des études de l'IEA.

Imagions que les 18 écoles reprises dans le *Tableau 23* appartiennent à la même strate explicite ? Dans ce cas, le facteur d'ajustement sera égal à la somme des produits $W_SCH_INIT6 * SIZE6$ pour les 18 écoles, divisé par la somme des produits $W_SCH_INIT6 * SIZE6$ pour les seules écoles participantes, soit $\frac{9077,8}{8037,5} = 1,1294$.

Les deux dernières colonnes du *Tableau 23* présentent les poids ajustés, ainsi que les produits poids ajustés*SIZE 6. La somme des valeurs de la dernière colonne s'élève pour les 16 écoles participantes à 9077,8, valeur parfaitement identique à ce qu'on observe sur les 18 écoles avant ajustement.

Plus les écoles seront groupées en fonction de critères qui corréleront avec la performance, plus le biais potentiel imputable à la non-réponse des écoles sera réduit. Les variables de stratification jouent donc un double rôle méthodologique, à savoir la réduction de la variance d'échantillonnage et la réduction du biais potentiel dû à la non-réponse des écoles. Leur importance mérite que les responsables internationaux les choisissent avec discernement.

4.5. Pondération classe et pondération élèves

Pour rappel, au sein des écoles échantillonnées, les administrateurs de tests doivent, dans un premier temps, sélectionner une classe parmi l'ensemble des classes de 6^e année du primaire ou de 9^e/10^e année pour le collège selon une procédure aléatoire et simple, et au sein de la classe, sélectionner 28 élèves (resp. 30 élèves au collège) parmi l'ensemble des élèves inscrits, qu'ils soient ou non présents le premier jour de l'administration des tests et questionnaires.

Dans le fichier Liste des écoles figure le nombre de classes de 6^e année (NB_CLASSES6). Comme une seule classe est choisie, le poids de la classe au sein de l'école se calcule comme suit :

$$W_CLA6 = \frac{1}{\frac{1}{NB_CLASSES6}} = NB_CLASSES6$$

Si l'école ne dispose que d'une seule classe, le poids de la classe au sein de l'école sera donc égal à 1.

Il ne reste plus, à ce stade que trois opérations pour obtenir le poids final de l'élève, à savoir :

1. calculer le poids initial de l'élève au sein de l'école ;
2. ajuster ce poids initial pour les élèves absents ;
3. combiner les différentes composantes pondérales pour obtenir le poids final de l'élève.

Pour rappel, les procédures de consistance entre les fichiers (i) *Liste des écoles*, (ii) *Formulaires de Suivi des élèves* et les fichiers de données (iii) *Tests* et (iv) *Questionnaire* ont été implémentées avec rigueur et, à ce stade des opérations, il ne peut plus exister aucune incohérence. Dans ces conditions, la poursuite des calculs liés à la pondération peut se poursuivre au niveau écoles, sans qu'il soit nécessaire d'intégrer les fichiers élèves.

La pondération initiale des élèves au sein de leur classe est donc égale à :

$$W_ELS_INIT6 = \frac{1}{\frac{NB_SAMP6}{NB_ELS6}} = \frac{NB_ELS6}{NB_SAMP6}$$

Il importe également à ce niveau de procéder à un ajustement pondéral pour la non-réponse des élèves. Le fichier *Liste des écoles*, fournit, par école :

1. le nombre d'élèves échantillonnés (NB_SAMP6 qui devrait, dans la majorité des cas, être égal à 28) ;
2. le nombre d'élèves qui ont été exclus (NB_EX6) ;

3. le nombre d'élèves absents (NB_AB6). Pour rappel, un élève absent est un élève échantillonné et qui n'a participé à aucune session d'évaluation ;
4. le nombre d'élèves présents (NB_PRE6). Pour rappel, un élève présent est un élève échantillonné qui a participé à une session d'évaluation au moins.

Le facteur d'ajustement pour la non-participation des élèves sera donc égal à :

$$ADJ_ELS_i = \frac{NB_PRE6 + NB_AB6}{NB_PRE6}$$

Les élèves exclus pour une des raisons susmentionnées n'interviennent pas, au même titre que les écoles exclues ne contribuent pas à l'ajustement pondéral pour la non-réponse des écoles.

Le poids final de l'élève comprend cinq composantes :

1. le poids initial de l'école ;
2. l'ajustement pour la non-réponse des écoles ;
3. le poids de la classe au sein de l'école ;
4. le poids initial de l'élève au sein de sa classe ;
5. l'ajustement pour la non-réponse des écoles.

Le poids final de l'élève est donc égal à :

$$W_ELS_FIN6 = W_SCH_ADJ_i * W_CLA6 * W_ELS_INIT6 * ADJ_ELS_i$$

Au sein d'une même classe, les élèves auront donc exactement le même poids.

4.6. Vérification des pondérations

Comme toute transformation de données, il convient de vérifier que les opérations ont été exécutées sans qu'une erreur se soit glissée dans la procédure. Il est donc vivement conseillé de réaliser plusieurs vérifications, parmi lesquelles on peut citer :

1. au niveau école :
 - tous les poids ont une valeur supérieur ou égale à 1 ;
 - pour une école donnée, le poids initial élèves au sein de leur école (W_ELS_INIT6) multiplié par la taille de l'échantillon au sein de cette école est égale au nombre d'élèves en 6^e, soit NB_ELS6 ;
 - le facteur d'ajustement pour la non-réponse des élèves doit être supérieur ou égal à un. De plus, ce facteur, multiplié par le nombre d'élèves présents correspond au nombre d'élèves absents plus le nombre d'élèves présents ;
2. au niveau de la strate explicite :
 - comparer la somme des pondérations élèves (soit au départ du fichier *Élèves*, soit au départ du fichier *Écoles* en multipliant le poids final élèves par le nombre de participants) et le comparer avec le nombre d'élèves total de sixième dans la *Base de sondage*. Si la différence dépasse 10 %, investiguer pour en déterminer l'origine ;
 - calculer le coefficient de variation du poids final de l'élève. Si ce coefficient est trop élevé, explorer les données pour en connaître les raisons.

4.7. Calcul du taux d'exclusion et de non-réponse des élèves

Les taux d'exclusion et de non-réponse des élèves peuvent se calculer au départ d'un fichier *Écoles*, pour autant qu'il ne comporte que les écoles participantes.

Le taux d'exclusion non pondéré est égal à :

$$T_EX_NP_ELS = \frac{\sum_{i=1} NB_EX6_i}{\sum_{i=1} (NB_PRE6_i + NB_AB6_i + NB_EX6_i)}$$

De même, le taux pondéré d'exclusion est égal à :

$$T_EX_P_ELS = \frac{\sum_{i=1} W_ELS_INIT6_i * NB_EX6_i}{\sum_{i=1} W_ELS_INIT6_i * (NB_PRE6_i + NB_AB6_i + NB_EX6_i)}$$

avec :

$$W_ELS_INIT6 = W_SCH_ADJ_i * W_CLA6 * W_ELS_INIT6$$

À l'inverse du poids final, ce poids initial de l'élève n'a pas été ajusté pour la non-réponse des élèves au sein des écoles.

De manière analogue, le taux non pondéré de participation des élèves est égal à :

$$T_R_NP_ELS = \frac{\sum_{i=1} NB_AB6_i}{\sum_{i=1} (NB_PRE6_i + NB_AB6_i)}$$

$$T_R_P_ELS = \frac{\sum_{i=1} W_ELS_INIT6_i * NB_AB6_i}{\sum_{i=1} W_ELS_INIT6_i * (NB_PRE6_i + NB_AB6_i)}$$

Le calcul des taux de réponse des élèves ne fait pas intervenir les élèves qui ont été exclus.

4.8. Pondération des données de 2^e année

La pondération des données de deuxième année suit globalement la même procédure que la pondération des données de sixième. La pondération initiale de l'école sera reprise de la pondération des données de sixième, et sera ensuite modifiée pour tenir compte du sous-échantillonnage des écoles.

Cet ajustement pour le sous-échantillonnage sera appliqué au niveau de la strate explicite. Ce facteur d'ajustement sera égal au rapport entre le nombre d'écoles échantillonnées pour la sixième année et le nombre d'écoles échantillonnées pour la deuxième année au sein de la strate explicite.

Ce poids final comprend donc cinq composantes :

1. le poids initial de l'école pour les données de 6^e année ;
2. l'ajustement pour le sous-échantillon ;
3. l'ajustement pour la non-réponse des écoles ;
4. le poids de la classe au sein de l'école ;
5. le poids initial de l'élève au sein de sa classe ;
6. l'ajustement pour la non-réponse des élèves.

Il se pourrait qu'une école soit exclue pour un niveau d'études donné, et pas pour l'autre. Ainsi, une école pourrait ne contenir que des élèves de sixième ou de deuxième ou inversement. Pour ces raisons entre autres, il est fortement conseillé de pondérer les données de deuxième indépendamment de la pondération des données de sixième.



1

2

3

4

5

Chapitre 5

Estimation de la variance d'échantillonnage



5.1. Introduction

Lorsque l'on collecte des données dans le cadre d'une étude internationale à large échelle à partir d'un échantillon d'individus, il existe des milliers d'échantillons de taille n possibles pour une même population et chacun de ces échantillons ne produit pas nécessairement les mêmes estimations de la population. L'estimation de statistiques de population à partir d'un échantillon d'individus comporte donc une certaine incertitude, ou risque d'erreur, liée à la variabilité qu'il existe entre les différents échantillons possibles (Neyman, 1934 ; Ardilly, 2006). La variance d'échantillonnage est donc une quantification de l'incertitude due à l'échantillonnage.

Ce chapitre décrit les procédures statistiques utilisées pour calculer la variance d'échantillonnage et sa racine carrée, l'erreur-type. Plus spécifiquement, il est discuté des méthodes d'estimation des variances d'échantillonnage par réplication des poids, lorsqu'une population est échantillonnée selon un plan d'échantillonnage complexe (Ardilly, 2006 ; Davison & Sardy, 2006). Premièrement, les concepts de variance d'échantillonnage et d'erreur-type sont introduits au travers d'un exemple fictif d'échantillonnage aléatoire et simple. Ensuite, le calcul d'erreur-type est envisagé dans le cas d'un échantillonnage à deux degrés. Enfin, les méthodes de réplication pour l'estimation des variances d'échantillonnage sont introduites dans le cas d'échantillons aléatoires et simples et dans le cas d'échantillons à deux degrés.

5.2. Variance d'échantillonnage d'échantillons aléatoires et simples

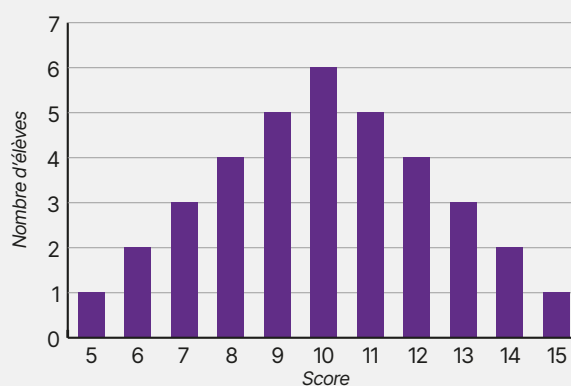
Supposons qu'un enseignant décide de mettre en place la pédagogie de maîtrise dans sa classe. Cette stratégie pédagogique nécessite que chaque unité d'apprentissage (ou leçon) soit suivie par une évaluation des élèves. L'enseignant a une classe de 36 élèves et réalise très rapidement que corriger à chaque évaluation les copies de tous les élèves lui prendrait beaucoup trop de temps. Il décide alors de sélectionner à chaque fois un échantillon de tests pour déterminer si la matière enseignée a été assimilée par les élèves et ainsi décider s'il peut poursuivre son enseignement ou s'il doit proposer une remédiation (Bloom, 1979).

Cependant, la sélection d'un échantillon aléatoire et simple de quelques tests peut conduire à sélectionner des élèves performants seulement ou à l'inverse, uniquement des élèves peu performants. Ces deux situations, peu probables, mais qui conduiraient à des décisions différentes, montrent néanmoins que le tirage d'un échantillon aléatoire génère toujours une certaine incertitude sur les résultats.

Dans notre exemple, l'enseignant est conscient de l'incertitude due à l'échantillonnage des tests. En vue de l'estimer, il décide, pour la première leçon, de corriger les tests des 36 élèves et d'analyser les résultats.

La Figure 7 présente la distribution des résultats obtenus par les 36 élèves : un élève a obtenu un score de 5, deux élèves obtiennent un score de 6... et un élève seulement obtient un score de 15.

FIGURE 7 : Distribution des résultats des 36 élèves



Source : Auteurs

La distribution des scores, comme l'illustre la *Figure 7*, s'apparente à une distribution normale. La moyenne de la population et la variance de la population sont respectivement égales à :

$$\mu = \frac{1}{N} \sum_{i=1}^N x_i = \frac{(5+6+6+7+\dots+14+14+15)}{36} = \frac{360}{36} = 10$$

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2 = \frac{[(5-10)^2 + (6-10)^2 + \dots + (14-10)^2 + (15-10)^2]}{36} = \frac{240}{36} = 5.8333$$

L'écart-type est alors égal à :

$$\sigma = \sqrt{\sigma^2} = \sqrt{5.8333} = 2.415$$

Pour gagner du temps, l'enseignant décide qu'il sélectionnera aléatoirement un échantillon de deux élèves après chaque leçon pour estimer l'assimilation de la matière par les élèves de sa classe. Le nombre d'échantillons de deux élèves issus d'une population de 36 élèves est égal à :

$$C_{36}^2 = \frac{36!}{(36-2)!2!} = 630$$

Le *Tableau 24* décrit les 630 échantillons de deux élèves possibles et leur moyenne au départ du premier test corrigé pour la population entière de la classe. Par exemple, il y a deux échantillons possibles qui génèrent une moyenne de population estimée à 5,5. Ces deux échantillons sont (i) l'élève dont le score est égal à 5 et le premier élève dont le score est égal à 6 et (ii) l'élève dont le score est égal à 5 et le second élève dont le score est égal à 6. Similairement, il y a deux manières de sélectionner un échantillon qui aurait une moyenne égale à 6 : les deux élèves qui ont obtenu un score 6 (un seul échantillon de ce type) ou l'élève avec le score de 5 et un des 3 élèves avec un score 7 (avec 3 échantillons de ce type puisqu'il y a un seul élève qui obtient le score 5 et 3 élèves qui obtiennent le score 7). Le *Tableau 24* montre qu'il y a deux échantillons possibles avec une moyenne de 5,5, quatre échantillons possibles avec une moyenne de 6, dix échantillons possibles avec une moyenne de 6,5, etc.

TABEAU 24 : Description des 630 échantillons de deux élèves possibles à partir d'une population de 36 élèves

Moyenne de l'échantillon	Scores des 2 élèves sélectionnés	Nombre de combinaisons de ces 2 scores	Nombre d'échantillons possibles
5.5	5 et 6	2	2
6.0	6 et 6	1	4
	5 et 7	3	
6.5	5 et 8	4	10
	6 et 7	6	
7.0	5 et 9	8	16
	6 et 8	3	
	7 et 7	5	
7.5	5 et 10	6	28
	6 et 9	10	
	7 et 8	12	
8.0	5 et 11	15	38
	6 et 10	6	
	7 et 9	5	
	8 et 8	12	

Moyenne de l'échantillon	Scores des 2 élèves sélectionnés	Nombre de combinaisons de ces 2 scores	Nombre d'échantillons possibles
8.5	5 et 12	4	52
	6 et 11	10	
	7 et 10	18	
	8 et 9	20	
9.0	5 et 13	24	60
	6 et 12	10	
	7 et 11	3	
	8 et 10	8	
	9 et 9	15	
9.5	5 et 14	2	70
	6 et 13	6	
	7 et 12	12	
	8 et 11	20	
	9 et 10	30	
10.0	5 et 15	1	70
	6 et 14	4	
	7 et 13	9	
	8 et 12	16	
	9 et 11	25	
	10 et 10	15	
10.5	6 et 15	2	70
	7 et 14	6	
	8 et 13	12	
	9 et 12	20	
	10 et 11	30	
11.0	7 et 15	3	60
	8 et 14	8	
	9 et 13	15	
	10 et 12	24	
	11 et 11	10	
11.5	8 et 15	4	52
	9 et 14	10	
	10 et 13	18	
	11 et 12	20	
12.0	9 et 15	5	38
	10 et 14	12	
	11 et 13	15	
	12 et 12	6	



1

2

3

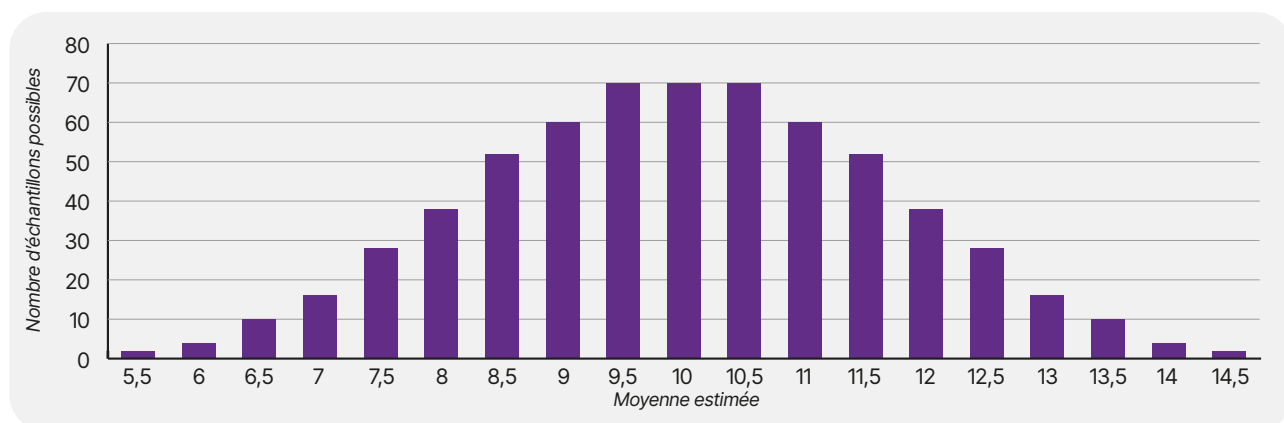
4

5

Moyenne de l'échantillon	Scores des 2 élèves sélectionnés	Nombre de combinaisons de ces 2 scores	Nombre d'échantillons possibles
12.5	10 et 15	6	28
	11 et 14	10	
	12 et 13	12	
13.0	11 et 15	5	16
	12 et 14	8	
	13 et 13	2	
13.5	12 et 15	4	10
	13 et 14	6	
14.0	13 et 15	3	4
	14 et 14	1	
14.5	14 et 15	2	2

La Figure 8 montre la distribution de la moyenne de tous les échantillons possibles de deux élèves tirés d'une population de 36 élèves.

FIGURE 8 : Distribution d'échantillonnage de la moyenne



Comme pour toute distribution, cette distribution des moyennes de tous les échantillons possibles peut être résumée par des indices de tendance centrale et par des indices de dispersion tels que la moyenne, notée à présent $\mu_{(\hat{\mu})}$ pour indiquer qu'il s'agit de la moyenne des estimations de la moyenne de la population et la variance $\sigma_{(\hat{\mu})}^2$ pour indiquer qu'il s'agit de la variance des estimations de moyenne :

$$\mu_{(\hat{\mu})} = [(2 * 5.5) + (4 * 6) + (10 * 6.5) + (16 * 7) + (28 * 7.5) + (38 * 8) + \dots + (2 * 14.5)] / 630 = 10$$

La moyenne de la moyenne de tous les échantillons possibles est égale à la moyenne de la population élève, c'est-à-dire 10. Ce résultat n'est pas une coïncidence mais une propriété fondamentale de la moyenne pour un échantillon aléatoire et simple : la moyenne des moyennes de tous les échantillons possibles est égale à la moyenne de la population. Dans un langage plus formel on dit que la moyenne d'un échantillon est un estimateur non biaisé de la moyenne de la population. On peut encore dire que l'espérance mathématique de la moyenne de l'échantillon est égale à la moyenne de la population.

Cependant, il est important de noter qu'il y a une grande variabilité autour de cette moyenne. Dans notre exemple, les moyennes des échantillons varient de 5,5 à 14,5. La variance de cette distribution, généralement appelée la variance d'échantillonnage de la moyenne, peut être calculée comme suit :

$$\sigma^2_{(\bar{\mu})} = \frac{1}{630} \sum_{i=1}^{630} (\hat{\mu}_i - \mu_{\hat{\mu}_i})^2$$

$$= [(5.5 - 10)^2 + (5.5 - 10)^2 + (6 - 10)^2 + \dots + (14.5 - 10)^2 + (14.5 - 10)^2] / 630 = 2.833$$

Sa racine carrée, appelée l'erreur-type, est égale à :

$$\sigma_{(\bar{\mu})} = \sqrt{\sigma^2_{(\bar{\mu})}} = \sqrt{2.833} = 1.68$$

Quelle information nous apporte l'erreur-type de la moyenne, ou spécifiquement dans notre exemple, que veut dire 1.68 ?

La distribution de la moyenne de tous les échantillons possibles suit approximativement une distribution normale. Dès lors, suivant les propriétés mathématiques de la distribution normale, on peut dire que :

- 68,2 % de tous les échantillons possibles ont une moyenne qui se situe entre -1 erreur-type et +1 erreur-type autour de la moyenne et ;
- 95,4 % de tous les échantillons possibles ont une moyenne qui se situe entre -2 erreurs-types et +2 erreurs-types autour de la moyenne.

Vérifions les propriétés mathématiques de la distribution normale à partir de la distribution d'échantillonnage de la moyenne dans notre exemple. La moyenne d'échantillonnage de la moyenne est ici égale à 10 et la racine carrée de sa variance, appelée dans ce cas l'« erreur-type », est égale à 1.68.

Combien d'échantillons ont une moyenne située entre $\mu_{(\bar{\mu})} - 1 \sigma_{(\bar{\mu})}$ et $\mu_{(\bar{\mu})} + 1 \sigma_{(\bar{\mu})}$ c'est-à-dire entre (10-1.68) et (10+1.68) ou encore entre 8.32 et 11.68 ?

Le *Tableau 25*, extrait du *Tableau 24*, montre qu'il y a 434 échantillons parmi les 630 possibles qui ont une moyenne comprise entre 8.32 et 11.68. Cela représente 68.9% de tous les échantillons. On peut également montrer qu'il y a 598 échantillons, soit 94.9% d'échantillons dont la moyenne est située entre $\mu_{(\bar{\mu})} - 2$ (6.64) et $\mu_{(\bar{\mu})} + 2$ (13.36).

TABLEAU 25 : Distribution de tous les échantillons possibles dont la moyenne se situe entre 8.32 et 11.68

Moyenne de l'échantillon	Nombre d'échantillons possibles	Pourcentage d'échantillons	Pourcentages cumulés d'échantillons
8.5	52	0.0825	0.0825
9.0	60	0.0952	0.1778
9.5	70	0.1111	0.2889
10.0	70	0.1111	0.4000
10.5	70	0.1111	0.5111
11.0	60	0.0952	0.6063
11.5	52	0.0825	0.6889
	434		

Dans cet exemple, pour calculer l'erreur-type sur la moyenne, la moyenne de tous les échantillons possibles a été calculée. Dans la réalité cependant, seule la moyenne d'un échantillon est connue. Nous allons montrer par la suite que la moyenne d'un seul échantillon est une donnée suffisante pour estimer la variance d'échantillonnage et donc l'erreur-type. Pour ce faire, il est important d'identifier les facteurs responsables de la variance d'échantillonnage d'un échantillon donné.

Dans la partie du document consacrée à l'échantillonnage des écoles, nous avons déjà évoqué que le premier facteur déterminant la variance d'échantillonnage était la taille de l'échantillon. Si l'enseignant de notre exemple, décide de sélectionner 4 élèves à la place de 2, alors la distribution d'échantillonnage de la moyenne s'étendra de 6 (les scores des 4 élèves les plus faibles étant 5, 6, 6 et 7) à 14 (les scores des 4 élèves les plus forts étant 13, 14, 14 et 15). Souvenons-nous qu'avec un échantillonnage de 2 élèves, la distribution d'échantillonnage allait de 5,5 à 14,5. Augmenter la taille de l'échantillon réduit donc la variance de la distribution.

Il y a 58 905 échantillons de 4 élèves possibles pris hors d'une population de 36 élèves. Le *Tableau 26* présente la distribution de tous les échantillons possibles de 4 élèves au départ d'une population de 36 élèves. Cette distribution a une moyenne de 10 et un écart-type, appelée erreur-type, de 1.115.

Cela prouve que la taille de l'échantillon n'affecte pas l'espérance mathématique de la moyenne de l'échantillon, mais la taille de l'échantillon réduit la variance de la distribution des moyennes des échantillons : plus la taille de l'échantillon est grande, plus la variance d'échantillonnage de la moyenne est faible. En d'autres termes, plus la taille de l'échantillon est grande, plus la précision des estimateurs est grande.

TABEAU 26 : Distribution de la moyenne de tous les échantillons de 4 élèves possibles à partir d'une population de 36 élèves

Nombre d'échantillons		Nombre d'échantillons	
Moyenne de l'échantillon	possibles	Moyenne de l'échantillon	possibles
6.00	3	10.25	4 830
6.25	10	10.5	4 553
6.5	33	10.75	4 082
6.75	74	11.00	3 531
7.00	159	11.25	2 896
7.25	292	11.5	2 288
7.5	510	11.75	1 700
7.75	804	12.00	1 213
8.00	1 213	12.25	804
8.25	1 700	12.5	510
8.5	2 288	12.75	292
8.75	2 896	13.00	159
9.00	3 531	13.25	74
9.25	4 082	13.5	33
9.5	4 553	13.75	10
9.75	4 830	14.00	3
10.00	4 949	-	-

Le second facteur qui contribue à la variance d'échantillonnage est la variance de la population elle-même. On peut facilement comprendre cette relation en faisant preuve de bon sens. Si une population est composée d'individus très semblables entre eux, c'est-à-dire une population avec peu de variance, les échantillons de taille n seront composés d'individus très semblables également et les échantillons différeront peu entre eux. La variance d'échantillonnage sera donc faible. À l'inverse, si une population est composée d'individus très différents, c'est-à-dire une population dont la variance est élevée, un échantillon de taille n issu de cette population pourra être très différent d'un second échantillon de même taille prélevé parmi cette même population. La variance d'échantillonnage sera plus élevée.

Si dans notre exemple, les résultats sont reportés sur un score total de 40 et non sur un score total de 20 (c'est-à-dire que les résultats de tous les élèves sont multipliés par deux), alors la moyenne des résultats de tous les élèves sera égale à 20, la variance sera égale à 23.333 (soit 4 fois 5.833) et l'écart-type sera égal à 4.83 (soit deux fois 2.415). Si un échantillon de deux élèves est prélevé, la variance d'échantillonnage sera égale à 11.333 (soit 4 fois 2.833) et l'erreur-type sur la moyenne sera égale à 3.3665 (soit deux fois 1.68).

L'erreur-type sur la moyenne est donc proportionnelle à la variance de la population et inversement proportionnel à la taille de l'échantillon. Il peut être montré que la variance d'échantillonnage de la moyenne, pour un échantillon sans remise au sein d'une population finie, est égale à :

$$\sigma_{(\bar{\mu})}^2 = \frac{\sigma^2}{n} \left(\frac{N-n}{N-1} \right)$$

Et l'erreur-type sur la moyenne de l'échantillon est égale à :

$$\sigma_{(\bar{\mu})} = \sqrt{\sigma_{(\bar{\mu})}^2} = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$$

Avec :

σ^2 = la variance de la population,

σ = l'écart-type de la population,

n = la taille de l'échantillon,

N = la taille de la population.

Cette formule appliquée à notre exemple initial, lorsque deux élèves sont sélectionnés, donne le résultat suivant :

$$\sigma_{(\bar{\mu})}^2 = \frac{\sigma^2}{n} \left(\frac{N-n}{N-1} \right) = \frac{5.833}{2} \left(\frac{36-2}{36-1} \right) = 2.8333$$

Plus la taille de la population augmente, plus le ratio $\frac{N-n}{N-1}$ tend vers 1 et la variance d'échantillonnage peut alors être

calculée sans tenir compte de la taille de la population N par la formule simplifiée suivante, valable dans le cadre d'un échantillon aléatoire et simple d'une population infinie :

$$\sigma_{(\bar{\mu})}^2 = \frac{\sigma^2}{n}$$

Cependant, en pratique, la variance de la population σ^2 n'est pas connue et est estimée à partir d'un échantillon. La variance d'échantillonnage de la moyenne estimée, comme toute estimation d'une moyenne, peut varier d'un échantillon à l'autre. En conséquence, à partir d'un échantillon, on ne pourra calculer qu'une estimation de la variance d'échantillonnage de la moyenne $\sigma_{(\bar{\mu})}^2$ (ou de toute autre statistique).

Dans la suite de ce document, les concepts de variance d'échantillonnage et d'estimation de la variance d'échantillonnage seront notés de la même façon pour simplifier le texte et les notations mathématiques. Cela veut dire que les estimations de la variance d'échantillonnage seront notées $\sigma_{(\bar{\mu})}^2$ sans utilisation du symbole $\wedge$ qui les différencie des valeurs vraies. Il faudra néanmoins comprendre qu'il s'agit bien d'estimations.

5.3. Variance d'échantillonnage d'un échantillon à plusieurs degrés

Les études internationales en sciences de l'éducation procèdent à un échantillonnage par degré, avec une première sélection d'écoles suivie d'une sélection de classes et/ou directement d'élèves (Kish, 1965).

Avec un échantillon à deux degrés, les élèves qui fréquentent la même école ne peuvent pas être considérés comme des observations indépendantes, à l'inverse d'un échantillon aléatoire et simple pour lequel les observations sont bien indépendantes. En effet, les élèves d'une même école ont généralement plus de caractéristiques communes que les élèves issus d'écoles différentes. Par exemple, ils partagent les mêmes ressources scolaires, peuvent avoir le même enseignant et donc sont instruits selon un même curriculum, etc. Les différences entre élèves issus de différentes écoles seront d'autant plus grandes que les écoles ne suivent pas toutes le même programme. Par exemple, on peut s'attendre à observer plus de différences entre les élèves issus d'une école d'enseignement professionnel et les élèves d'une école d'enseignement général qu'entre les élèves issus de deux écoles d'enseignement professionnel.

De plus, dans un pays, au sein d'une région et d'une ville, les personnes ont tendance à vivre dans certains quartiers selon leurs ressources financières. Comme les enfants fréquentent généralement une école proche de leur habitation, il est probable que les élèves fréquentant une même école soient issus de familles de niveaux socio-économiques proches. Par contre dans les zones rurales, les écoles étant moins nombreuses, le facteur socio-économique n'intervient certainement pas comme facteur d'homogénéisation des écoles.

Pour les raisons développées ci-dessus, un échantillon aléatoire et simple de 4000 élèves sera plus représentatif de la diversité de la population qu'un échantillon de 100 écoles avec 40 élèves évalués dans chacune d'elles. En conséquence, l'incertitude associée à l'estimation de tout paramètre de population – c'est-à-dire l'erreur-type – sera plus grande pour un échantillon à deux degrés que pour un échantillon aléatoire et simple de même taille.

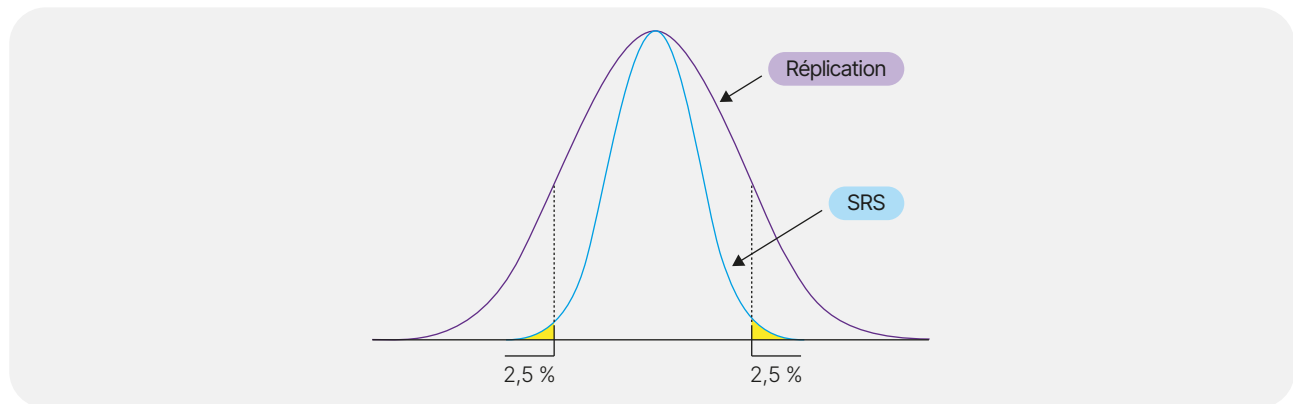
L'accroissement de l'incertitude pour un échantillon à deux degrés est directement proportionnel aux différences qui existent entre les unités de premier niveau d'échantillonnage (en anglais *Primary Sampling Unit* ou PSU) (Ardilly, 2006), c'est-à-dire les écoles. On peut le montrer au travers des deux cas extrêmes et fictifs ci-dessous :

- soit un système éducatif fictif où tous les élèves sont assignés de manière complètement aléatoire à une école. Sélectionner aléatoirement 100 écoles et ensuite 40 élèves dans chacune d'elles sera tout à fait similaire d'un point de vue statistique que sélectionner directement et aléatoirement 4 000 élèves puisqu'il n'y a pas de différences entre les écoles. Dans ce cas, l'incertitude associée à l'estimation des paramètres de population sera égale à celle obtenue pour un échantillon aléatoire et simple de 4 000 élèves ;
- soit un système éducatif fictif où les écoles sont différentes mais où les élèves d'une même école sont parfaitement identiques. Puisque tous les élèves d'une école sont identiques, observer 40 élèves dans une école n'apportera aucune information supplémentaire que d'observer un seul élève. Dans ce cas, si l'on sélectionne aléatoirement 100 écoles et puis 40 élèves par école sélectionnée, la taille effective de l'échantillon sera égale à 100. En conséquence, l'incertitude associée à l'estimation des paramètres de population, l'erreur-type, sera égale à celle obtenue pour un échantillon aléatoire et simple de 100 élèves et sera donc substantiellement plus élevée que dans l'exemple extrême contraire.

Bien sûr aucun système éducatif au monde ne peut être assimilé à ces deux cas extrêmes. Cependant, dans certains systèmes éducatifs, les différences entre écoles sont très petites, du moins en ce qui concerne les variables mesurées par l'étude et notamment les performances en lecture et en mathématiques. Dans d'autres systèmes éducatifs par contre, les différences entre écoles sont plus importantes. Les études antérieures du PASEC ont par exemple montré que les différences entre écoles sont assez modérées à Madagascar alors qu'elles sont plus élevées en Mauritanie ou au Cameroun.

Au travers de ces deux exemples fictifs, il apparaît clairement que le plan d'échantillonnage à plusieurs niveaux doit être pris en considération pour l'estimation de la variance d'échantillonnage. Estimer la variance d'échantillonnage d'un échantillon à deux niveaux comme s'il s'agissait d'un échantillon aléatoire et simple revient à sous-estimer substantiellement cette variance d'échantillonnage et par conséquent les intervalles de confiance des paramètres de population estimés seront trop étroits (Srinath & Hidirolou, 1977). Un effet non significatif pourrait alors apparaître erronément comme significatif (erreur de type I). La Figure 9 illustre cette sous-estimation de l'erreur d'échantillonnage. Si la variance d'échantillonnage est calculée comme s'il s'agissait d'un échantillon aléatoire et simple (SRS), la distribution est nettement plus étroite que si la méthode utilisée tient compte des caractéristiques de l'échantillon à plusieurs niveaux (méthode des répliques décrites ci-dessous). La figure montre que le risque d'erreur de type I est important.

FIGURE 9 : Distribution d'échantillonnage pour l'estimation de la moyenne calculée comme pour un SRS d'écoles et distribution d'échantillonnage attendue pour un échantillon PPS d'écoles



L'erreur de type I commise sera d'autant plus grande que la variance entre écoles est grande. L'effet du plan d'échantillonnage (en anglais *sampling design effect* ou SDE) est une mesure de la sous-estimation de la variance d'échantillonnage lorsque celle-ci est calculée en considérant l'échantillon comme un échantillon SRS d'écoles. Cet effet d'échantillonnage est égal à (OECD, 2009) :

$$SDE = \frac{\sigma^2_{(\hat{\theta}_{rep})}}{\sigma^2_{(\hat{\theta}_{SRS})}}$$

Le *Tableau 27* montre l'ampleur de l'erreur de type I selon la valeur de l'effet du plan de sondage.

TABEAU 27 : Effet d'échantillonnage et erreur de type I

Effet d'échantillonnage	Erreur de Type 1	Effet d'échantillonnage	Erreur de Type 1
1.5	0.11	11.0	0.55
2.0	0.17	11.5	0.56
2.5	0.22	12.0	0.57
3.0	0.26	12.5	0.58
3.5	0.29	13.0	0.59
4.0	0.33	13.5	0.59
4.5	0.36	14.0	0.60
5.0	0.38	14.5	0.61
5.5	0.40	15.0	0.61
6.0	0.42	15.5	0.62
6.5	0.44	16.0	0.62
7.0	0.46	16.5	0.63
7.5	0.47	17.0	0.63
8.0	0.49	17.5	0.64
8.5	0.50	18.0	0.64
9.0	0.51	18.5	0.65
9.5	0.52	19.0	0.65
10.0	0.54	19.5	0.66
10.5	0.55	20.0	0.66

Le plan d'échantillonnage à deux niveaux le plus simple qu'on puisse imaginer est celui appliqué au départ d'une population infinie d'unités de premier niveau et une population infinie d'unités de second niveau (ou finies avec remise puisque un échantillon avec remise au sein d'une population finie revient à un échantillon sans remise au sein d'une population infinie). Les deux niveaux d'unités étant des populations infinies, les unités de premier niveau (PSUs) sont considérées comme étant de tailles égales. Dans ce cas, si un échantillon aléatoire et simple de PSUs est sélectionné et si, dans chaque unité sélectionnée, un échantillon aléatoire et simple d'unités de second niveau est également sélectionné, alors la variance d'échantillonnage de la moyenne sera égale à :

$$\sigma^2_{(\bar{\mu})} = \frac{\sigma^2_{\text{Between-PSU}}}{n_{\text{PSU}}} + \frac{\sigma^2_{\text{Within-PSU}}}{n_{\text{PSU}} n_{\text{within}}}$$

Les plans d'échantillonnage utilisés dans les grandes études des systèmes éducatifs sont généralement des plans à deux niveaux beaucoup plus complexes que celui décrit ci-dessus et la variance d'échantillonnage ne peut être calculée à l'aide de cette simple formule. En effet, le calcul devrait intégrer d'autres caractéristiques qui influent sur la variance d'échantillonnage, à savoir :

- les populations d'écoles et d'élèves sont des populations finies et non des populations infinies ;
- les écoles sont de tailles différentes ;
- les écoles ne sont pas sélectionnées par SRS mais selon avec une probabilité proportionnelle à la taille (PPS) selon une procédure systématique ;
- des variables de stratification explicite et implicite sont incluses...

Certaines de ces caractéristiques d'échantillonnage sont implémentées dans le but de réduire la variance d'échantillonnage. Ne pas en tenir compte dans le calcul par l'utilisation de la formule mathématique ci-dessus entraînerait par conséquent une surestimation de la variance d'échantillonnage.

Depuis l'étude IEA Reading Literacy de 1990, des méthodes de réplication ou de ré-échantillonnage sont utilisées pour estimer les variances d'échantillonnage. Ces méthodes impliquent de nombreuses étapes de calculs en série ; elles sont connues depuis la fin des années 50 mais n'ont pu être largement utilisées qu'à partir des années 90, avec l'apparition d'ordinateurs individuels suffisamment puissants.

Dans les paragraphes qui suivent, le principe des méthodes de réplication va d'abord être expliqué pour l'échantillon le plus simple qu'il soit, l'échantillon aléatoire et simple. Les méthodes de réplication seront ensuite décrites pour des plans d'échantillonnage plus complexes. Une attention particulière sera apportée à la méthode qui sera adoptée pour la prochaine enquête du PASEC.

5.4. Les méthodes de réplication pour les échantillons aléatoires et simples

Toutes les méthodes de réplication reposent sur un même processus en trois étapes :

1. la reproduction d'un nombre important de (sous-)échantillons fictifs au départ de l'échantillon réel ;
2. le calcul de la statistique pour l'échantillon entier et pour chacun des (sous-)échantillons répliqués ;
3. l'analyse de la variation du paramètre estimé pour l'échantillon entier et pour chaque sous-échantillon.

Il existe principalement deux types de méthodes de réplication pour échantillons aléatoires et simples : la méthode Jackknife (Quenouille, 1949) et la méthode Bootstrap (Bertail & Combris, 1997 ; Rao & Wu, 1988 ; Efron, 1979). Une des plus grandes différences entre ces deux méthodes est la procédure utilisée pour produire les réplifications, c'est-à-dire les sous-échantillons. Pour un échantillon de n unités, le Jackknife génère de manière systématique n sous-échantillons répliqués de $n-1$ unités. Le Bootstrap, par contre, génère de façon aléatoire un grand nombre d'échantillons répétés de n unités sélectionnées avec remplacement, chaque unité pouvant être repris plus d'une fois dans une réplication.

Dans le cadre de la prochaine étude du PASEC, tout comme l'étude PIRLS ou TIMSS, l'utilisation d'une méthode de réplication de type Jackknife est fortement conseillée. Dans les lignes qui suivent, le principe de réplication pour un échantillon aléatoire et simple est donc principalement expliqué suivant la méthode Jackknife ; ensuite, et pour information, le principe de réplication pour un échantillon SRS est également expliqué selon la méthode Bootstrap.

La méthode Jackknife pour un échantillon aléatoire et simple

Supposons un échantillon aléatoire et simple de 10 élèves. La méthode Jackknife va générer 10 sous-échantillons, ou répliques, de 9 élèves chacun, tels que décrits dans le *Tableau 28* (Quenouille, 1949 ; Tukey, 1958).

TABLEAU 28 : Les échantillons répliqués Jackknife et leur moyenne

Élève	1	2	3	4	5	6	7	8	9	10	Moyenne
Score	10	11	12	13	14	15	16	17	18	19	14.50
Réplique 1	0	1	1	1	1	1	1	1	1	1	15.00
Réplique 2	1	0	1	1	1	1	1	1	1	1	14.88
Réplique 3	1	1	0	1	1	1	1	1	1	1	14.77
Réplique 4	1	1	1	0	1	1	1	1	1	1	14.66
Réplique 5	1	1	1	1	0	1	1	1	1	1	14.55
Réplique 6	1	1	1	1	1	0	1	1	1	1	14.44
Réplique 7	1	1	1	1	1	1	0	1	1	1	14.33
Réplique 8	1	1	1	1	1	1	1	0	1	1	14.22
Réplique 9	1	1	1	1	1	1	1	1	0	1	14.11
Réplique 10	1	1	1	1	1	1	1	1	1	0	14.00

La moyenne de l'échantillon initial est égale à 14.5. Pour la première réplique, l'élève 1 (dont le score est égal à 10) est exclu du sous-échantillon et la moyenne des 9 élèves est égale à 15.0. Pour la deuxième réplique, l'élève 2 (dont le score est égal à 11) est exclu et la moyenne des 9 élèves inclus est égale à 14.88, etc.

L'estimation Jackknife de la variance d'échantillonnage de la moyenne est égale à :

$$\sigma_{jack}^2 = \frac{n-1}{n} \sum_{i=1}^n (\hat{\theta}_{(i)} - \hat{\theta})^2 \quad \text{Méthode de réplication Jackknife}$$

Où $\hat{\theta}_{(i)}$ représente la statistique estimée pour la réplique i et $\hat{\theta}$ représente la statistique estimée pour l'échantillon entier.

Dans notre exemple, la variance d'échantillonnage de la moyenne estimée par Jackknife est égale à :

$$\sigma_{(\hat{\mu})}^2 = \frac{9}{10} [(15.00 - 14.50)^2 + (14.88 - 14.50)^2 + \dots + (14.11 - 14.50)^2 + (14.00 - 14.50)^2]$$

$$\sigma_{(\hat{\mu})}^2 = \frac{9}{10} (1.018519) = 0.9167$$

Calculée par la formule mathématique, la variance d'échantillonnage de la moyenne est égale à :

$$\sigma_{(\hat{\mu})}^2 = \frac{\sigma^2}{n} \quad \text{Calcul mathématique}$$

Avec σ^2 , la variance de population estimée à partir de la variance de l'échantillon total, égale à :

$$\sigma^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \hat{\mu})^2 = \frac{1}{9} [(10 - 14.50)^2 + (11 - 14.50)^2 + \dots + (18 - 14.50)^2 + (19 - 14.50)^2] = 9.17$$

Mathématiquement, la variance d'échantillonnage de la moyenne est donc égale à :

$$\sigma_{(\hat{\mu})}^2 = \frac{\sigma^2}{n} = \frac{9.17}{10} = 0.917$$

Cet exemple montre donc que la variance d'échantillonnage de la moyenne estimée par la méthode Jackknife est identique à celle estimée par la formule mathématique. Ce n'est pas une coïncidence, et cette égalité a été démontrée mathématiquement par Rust et Rao (1996).

La méthode Jackknife peut également être appliquée au calcul de la variance d'échantillonnage d'autres statistiques que la moyenne. Les *Tableaux 29 et 30* montrent à partir du même échantillon fictif de 10 élèves, l'application de la méthode pour le calcul de la variance d'échantillonnage du coefficient de régression. De façon identique, la procédure consiste à calculer le coefficient de régression de l'échantillon total et ceux des 10 sous-échantillons. La comparaison du coefficient de régression de l'échantillon total avec les coefficients de régression de chacun des 10 sous-échantillons permettra l'estimation de la variance d'échantillonnage de cette statistique.

TABLEAU 29 : Valeur des variables X et Y pour un échantillon de 10 élèves

Élève	1	2	3	4	5	6	7	8	9	10
Variable X	10	11	12	13	14	15	16	17	18	19
Variable Y	10	13	14	19	11	12	16	17	18	15

TABLEAU 30 : Coefficients de régression pour l'échantillon total et les 10 répliques

	Coefficient de régression
Échantillon total	0.53
Réplique 1	0.35
Réplique 2	0.55
Réplique 3	0.56
Réplique 4	0.64
Réplique 5	0.51
Réplique 6	0.55
Réplique 7	0.51
Réplique 8	0.48
Réplique 9	0.43
Réplique 10	0.68

Selon la méthode Jackknife, la variance d'échantillonnage du coefficient de régression calculée à partir des données du *Tableau 30* est égale à :

$$\sigma_{jack}^2 = \frac{n-1}{n} \sum_{i=1}^n (\hat{\theta}_{(i)} - \hat{\theta})^2$$

$$\sigma_{jack}^2 = \frac{9}{10} [(0.35 - 0.53)^2 + (0.55 - 0.53)^2 + \dots + (0.68 - 0.53)^2] = 0.07$$

On peut à nouveau montrer que ce résultat obtenu par Jackknife est identique à celui que l'on obtiendrait si la variance d'échantillonnage du coefficient de régression était calculée par la formule mathématique habituelle.

La méthode Bootstrap pour un échantillon aléatoire et simple

Supposons à nouveau un échantillon aléatoire et simple de 10 élèves. La méthode Bootstrap va générer de façon aléatoire un très grand nombre d'échantillons parmi tous les échantillons possibles de 10 élèves avec répétition (soit $\Gamma_{10}^{10} = C_{19}^{10} = \frac{19!}{10! \cdot 9!} = 92378$ échantillons possibles). Le *Tableau 31* ci-dessous présente globalement la structure de la matrice

d'échantillons répliqués selon Bootstrap (Efron, 1979).

TABLEAU 31 : Exemple de matrice d'échantillons Bootstrap

Élève	1	2	3	4	5	6	7	8	9	10	Moyenne
Score	10	11	12	13	14	15	16	17	18	19	14.50
Structure 1	1	1	1	1	1	1	1	1	1	1	14.50
Structure 2	2	1	1	1	1	1	1	1	1	0	De 13.6 à 15.4
Structure 3	2	2	1	1	1	1	1	1	0	0	De 12.9 à 16.1
Structure	6	1	1	1	1	0	0	0	0	0	
Structure	7	1	1	1	0	0	0	0	0	0	
Structure	8	1	1	0	0	0	0	0	0	0	
Structure	9	1	0	0	0	0	0	0	0	0	
Structure	10	0	0	0	0	0	0	0	0	0	De 10 à 19

90 répliqués répondent à la structure 2 et le score moyen de ces échantillons va de 13.6 (si l'élève 10 n'est pas comptabilisé et que l'élève 1 est pris deux fois) à 15.4 (si l'élève 1 n'est pas sélectionné et l'élève 10 est pris deux fois). Dix échantillons répondent à la dernière structure et leur moyenne va de 10 si l'élève 1 est répété 10 fois à 19 si l'élève 10 est répété 10 fois.

Pour un SRS, l'estimation Bootstrap de la variance d'échantillonnage de la moyenne est égale à :

$$\sigma_{boot}^2 = \frac{1}{G-1} \sum_{i=1}^n (\hat{\theta}_{(i)} - \hat{\theta}_{(i)})^2 \quad \text{Méthode de répliqués Bootstrap}$$

G étant le nombre de PSUs (écoles) et n le nombre de répliqués.

5.5. Méthodes de répliqués pour les échantillons à deux niveaux

Il y a trois types de méthodes de répliqués pour les échantillons à deux niveaux (Wolter, 2007 ; Rust & Rao, 1996) :

1. le Jackknife avec deux variantes : une pour des échantillons non stratifiés (Jackknife 1) et une seconde pour des échantillons stratifiés (Jackknife 2) ;
2. la méthode des répliqués équilibrés répétées (en anglais, *Balanced Repeated Replication* ou BRR, traduit aussi demi-échantillonnage équilibré) et sa variante, la méthode de Fay qui est utilisée dans PISA et dans Talis ;
3. le Bootstrap.

La prochaine enquête du PASEC utilisera le Jackknife 2.

La méthode Jackknife pour un plan d'échantillonnage non stratifié à deux niveaux

Soit un échantillon de 10 écoles et parmi ces écoles, un échantillon aléatoire et simple d'élèves. La méthode du Jackknife pour un échantillon non stratifié consiste à générer 10 réplifications de 9 écoles. Comme présenté dans le *Tableau 32*, chaque école est retirée d'un seul sous-échantillon, selon une procédure systématique.

Pour la première réplification notée R1, l'école 1 est retirée. Le poids des écoles restantes de la réplification 1 est ajusté par un facteur $\frac{G}{G-1}$, G étant le nombre de PSUs (écoles) mais aussi le nombre de répliques dans l'échantillon. Dans notre exemple,

ce facteur vaut 10/9, soit 1.11.

Ce facteur d'ajustement sera ensuite appliqué pour calculer le poids élève de la réplification par combinaison du poids école et du poids à l'intérieur de l'école pour cette même réplification.

TABLEAU 32 : Réplication Jackknife pour un échantillonnage non stratifié à deux niveaux

Réplique	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10
École 1	0.00	1.11	1.11	1.11	1.11	1.11	1.11	1.11	1.11	1.11
École 2	1.11	0.00	1.11	1.11	1.11	1.11	1.11	1.11	1.11	1.11
École 3	1.11	1.11	0.00	1.11	1.11	1.11	1.11	1.11	1.11	1.11
École 4	1.11	1.11	1.11	0.00	1.11	1.11	1.11	1.11	1.11	1.11
École 5	1.11	1.11	1.11	1.11	0.00	1.11	1.11	1.11	1.11	1.11
École 6	1.11	1.11	1.11	1.11	1.11	0.00	1.11	1.11	1.11	1.11
École 7	1.11	1.11	1.11	1.11	1.11	1.11	0.00	1.11	1.11	1.11
École 8	1.11	1.11	1.11	1.11	1.11	1.11	1.11	0.00	1.11	1.11
École 9	1.11	1.11	1.11	1.11	1.11	1.11	1.11	1.11	0.00	1.11
École 10	1.11	1.11	1.11	1.11	1.11	1.11	1.11	1.11	1.11	0.00

Toute statistique est alors calculée pour l'échantillon entier et pour chacune des réplifications. La variance d'échantillonnage est calculée en comparant l'estimation pour chaque réplique à l'estimation pour l'échantillon entier, selon la formule suivante :

$$\sigma^2_{(\hat{\theta})} = \frac{G-1}{G} \sum_{i=1}^G (\hat{\theta}_{(i)} - \hat{\theta})^2 \quad \text{Méthode de réplification Jackknife}$$

Cette formule est quasi identique à celle utilisée pour un échantillon aléatoire et simple. Seul le nombre de répliques G , correspondant au nombre de PSUs (écoles), remplace le nombre de réplifications n correspondant au nombre d'unités dans un échantillon aléatoire et simple.

Dans ce cas également, on pourrait montrer que la variance d'échantillonnage de la moyenne obtenue par la méthode de Jackknife est identique à celle qui serait obtenue mathématiquement par la formule déjà présentée précédemment pour le calcul de la variance d'échantillonnage d'un échantillon simple à deux niveaux :

$$\sigma^2_{(\hat{\mu})} = \frac{\sigma^2_{\text{Between_PSU}}}{n_{\text{PSU}}} + \frac{\sigma^2_{\text{Within_PSU}}}{n_{\text{PSU}} n_{\text{within}}} \quad \text{Calcul mathématique}$$

La méthode Jackknife pour un plan d'échantillonnage stratifié à deux niveaux (JK2)

Nous avons déjà mentionné dans la partie III que le plan d'échantillonnage doit être conçu pour éviter tout biais dans la procédure de sélection. Il doit également assurer un maximum de précision dans l'estimation des paramètres de population, sur base des moyens financiers disponibles (Johnson & Rust, 1992 ; Martin et al., 2003).

Pour augmenter la précision, c'est-à-dire minimiser la variance d'échantillonnage, sans augmenter la taille de l'échantillon, les études internationales ont recouru à plusieurs processus intégrés au plan d'échantillonnage :

- les écoles sont sélectionnées selon une procédure systématique, avec une probabilité proportionnelle à leur taille et des échantillons équivalents d'élèves sont échantillonnés dans les écoles sélectionnées. Cette procédure garantit que la probabilité finale de sélection des élèves ne varie pas beaucoup ;
- des variables de stratification sont utilisées (les variables « région », « type d'école » et « urbanisation » étant les 3 variables suggérées par le PASEC). La réduction de la variance d'échantillonnage induite par la stratification sera d'autant plus grande que les variables de stratification sont liées à la performance élève mesurée.

La méthode Jackknife pour échantillonnage stratifié à plusieurs niveaux prend en considération ces aspects du plan d'échantillonnage et permet ainsi d'estimer à sa juste valeur la variance d'échantillonnage.

Supposons que la *Base de sondage* d'une population est divisée en deux strates : les écoles rurales et les écoles urbaines. De plus, à l'intérieur de chacune des deux strates, les écoles sont triées par taille. Dans chaque strate, 10 écoles sont sélectionnées selon une procédure systématique avec une probabilité proportionnelle à leur taille.

La méthode Jackknife 2 pour un échantillon stratifié à plusieurs niveaux consiste à grouper, au sein de chaque strate, les unités premièrement échantillonnées par paires selon l'ordre dans lequel elles ont été sélectionnées. Dans le cadre d'un échantillon de type PIRLS où des écoles sont sélectionnées et puis une classe dans chacune des écoles participantes, on formera des paires d'écoles. Ainsi chaque école est appariée à une école qui lui est similaire.

Le *Tableau 33* montre comment les répliques sont générées par cette méthode. Les écoles 1 à 10 sont les écoles échantillonnées dans la strate « rurale » et les écoles 11 à 20 sont les écoles échantillonnées dans la strate « urbaine ». Dans chaque strate, il y a donc 5 paires d'écoles ou pseudo-strates (également appelées « strate de variance » ou encore zone de Jackknife).

TABLEAU 33 : Les réplifications Jackknife pour un échantillon stratifié à deux niveaux

	Pseudo- strate	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10
École 1	1	2	1	1	1	1	1	1	1	1	1
École 2	1	0	1	1	1	1	1	1	1	1	1
École 3	2	1	0	1	1	1	1	1	1	1	1
École 4	2	1	2	1	1	1	1	1	1	1	1
École 5	3	1	1	2	1	1	1	1	1	1	1
École 6	3	1	1	0	1	1	1	1	1	1	1
École 7	4	1	1	1	0	1	1	1	1	1	1
École 8	4	1	1	1	2	1	1	1	1	1	1
École 9	5	1	1	1	1	2	1	1	1	1	1
École 10	5	1	1	1	1	0	1	1	1	1	1
École 11	6	1	1	1	1	1	2	1	1	1	1
École 12	6	1	1	1	1	1	0	1	1	1	1
École 13	7	1	1	1	1	1	1	0	1	1	1
École 14	7	1	1	1	1	1	1	2	1	1	1
École 15	8	1	1	1	1	1	1	1	0	1	1
École 16	8	1	1	1	1	1	1	1	2	1	1
École 17	9	1	1	1	1	1	1	1	1	0	1
École 18	9	1	1	1	1	1	1	1	1	2	1
École 19	10	1	1	1	1	1	1	1	1	1	2
École 20	10	1	1	1	1	1	1	1	1	1	0

La méthode générera autant de réplifications dans la strate qu'il y a de paires ou pseudo-strates. Dans notre exemple, 10 réplifications sont produites. Chaque réplification est créée en retirant aléatoirement une école pour une pseudo-strate particulière, le poids de l'école restante dans la pseudo-strate étant alors doublé. Pour la réplification 1, nommée R_1 , l'école 2 a été aléatoirement retirée de la pseudo-strate 1 et le poids de l'école 1 a été doublé dans la pseudo-strate. Pour la réplification 2, nommée R_2 , l'école 3 a été aléatoirement retirée de la pseudo-strate 2 et le poids de l'école 4 a été doublé dans la pseudo-strate, etc.

À nouveau, la variance d'échantillonnage d'une statistique est calculée en comparant cette statistique estimée pour l'échantillon entier avec celles estimées pour chacune des réplifications, selon la formule suivante :

$$\sigma^2_{(\hat{\theta})} = \sum_{i=1}^G (\hat{\theta}_{(i)} - \hat{\theta})^2$$

Cette formule peut être utilisée quelle que soit la statistique estimée : moyenne, corrélation, pourcentages, totaux...

Cette méthode de réplification JK2 est la méthode généralement appliquée dans les enquêtes internationales menées par l'IEA (TIMSS, PIRLS, SITES...). C'est également cette méthode qui sera utilisée pour calculer les variances d'échantillonnage de la prochaine enquête du PASEC.

L'enquête PISA applique la méthode de Fay qui est une variante du Jackknife ou plus exactement de la méthode des répliquions équilibrées répétées (en anglais, *the Balanced Repeated Replication method* ou BRR). Alors que la méthode JK2 consiste à retirer seulement une école par répliquion, la méthode BRR procède en retirant une école par pseudo-strate et en doublant le poids de l'école restante au sein de chaque pseudo-strate. Chaque répliquion est donc un sous-échantillon constitué de la moitié des écoles de l'échantillon complet. Théoriquement, il existe une infinité de répliquions. Pour maximiser l'estimation de la variance d'échantillonnage, les répliquions sont conçues selon des matrices Hadamard. Le nombre de répliquion doit, selon cette méthode, être le plus petit multiple de 4 plus grand que le nombre de paires. Ainsi, avec 75 paires d'écoles, (pour un échantillon de 150 écoles), 76 répliquions seront constituées. Cette méthode n'est pas développée davantage dans ce document mais des informations complémentaires peuvent être trouvées dans le manuel d'analyses des données PISA (OECD, 2009).

Illustrons la méthode JK2, utilisée par le PASEC, dans un exemple concret. Soit 10 écoles, dont les scores moyens et les poids de chaque école dans les différentes répliquions sont présentés dans le *Tableau 34* ci-dessous.

TABLEAU 34 : Répliquions selon la méthode JK2

	Score moyen	Pseudo-strate	R1	R2	R3	R4	R5
École 1	100	1	2	1	1	1	1
École 2	110	1	0	1	1	1	1
École 3	120	2	1	0	1	1	1
École 4	130	2	1	2	1	1	1
École 5	140	3	1	1	2	1	1
École 6	145	3	1	1	0	1	1
École 7	160	4	1	1	1	0	1
École 8	180	4	1	1	1	2	1
École 9	190	5	1	1	1	1	2
École 10	195	5	1	1	1	1	0

La moyenne de l'échantillon complet est égale à 147. La moyenne de la première répliquion R1 est égale à :

$$\hat{\mu}_{R_1} = \frac{(100 \cdot 2 + 105 \cdot 0 + 120 \cdot 1 + \dots + 200 \cdot 1)}{10} = 146$$

De même, $\hat{\mu}_{R_2} = 148$, $\hat{\mu}_{R_3} = 146.5$, $\hat{\mu}_R = 149$ et $\hat{\mu}_{R_5} = 146.5$

La variance d'échantillonnage de la moyenne estimée selon la méthode JK2 est donc égale à :

$$\sigma_{(\hat{\theta})}^2 = \sum_{i=1}^G (\hat{\theta}_{(i)} - \hat{\theta})^2 = (148 - 147)^2 + (146.5 - 147)^2 + \dots + (146.5 - 147)^2 = 6.5$$

5.6. Procédures techniques d'estimation de la variance d'échantillonnage par JK2

La variance d'échantillonnage étant fonction de la taille de l'échantillon, les procédures décrites ci-dessous doivent être implémentées séparément pour l'échantillon de 6^e primaire et pour l'échantillon de 2^e primaire. La variance d'échantillonnage estimée pour la 2^e primaire sera supérieure à celle estimée pour la 6^e primaire, la taille d'échantillon étant deux fois plus petite.

La technique d'application de la méthode JK2 comprend :

- la construction des zones de Jackknife ou pseudo-strates ;
- la génération des répliques ;
- le calcul de la variance d'échantillonnage.

Pour construire les zones de Jackknife, les écoles participantes sont d'abord triées selon l'ordre dans lequel elles ont été sélectionnées dans l'échantillon, cet ordre induisant notamment un tri automatique selon les variables de stratification explicite et implicite. Dans cette liste ordonnée, une école de remplacement participante prend la place de l'école originale qu'elle remplace. Une école sélectionnée non-participante et non remplacée n'est pas reprise dans la base de données.

Au sein de chaque strate explicite, des couples d'écoles sont constitués : dans la liste ordonnée, l'école 1 et l'école 2 sont assignées à la première pseudo-strate, l'école 3 et l'école 4 sont assignées à la deuxième pseudo-strate, etc. Si la strate explicite contient un nombre impair d'écoles, les élèves de la dernière école de la strate sont aléatoirement répartis en deux groupes pour constituer deux « pseudo-écoles ».

L'échantillon principal du primaire étant constitué de 200 écoles, 100 pseudo-strates sont attendues pour l'échantillon de 6^e primaire et 50 pseudo-strates pour celui de 2^e primaire. Celui du collège est constitué de 180 établissements avec 90 pseudo-strates.

À primaire, si pour un pays donné, le plan d'échantillonnage conduit à former plus de pseudo-strates qu'attendu, les pseudo-strates au-delà de la pseudo-strate 100 seront fusionnées pour ne pas dépasser le nombre de pseudo-strates attendu. Ainsi, la pseudo-strate 101 sera fusionnée avec la pseudo-strate 1, la 102 avec la 2 et ainsi de suite. C'est égal au collège où les pseudo-strates au-delà de la pseudo-strate 90 seront fusionnées. La pseudo-strate 91 sera fusionnée avec la pseudo-strate 1, la 92 avec la 2 et ainsi de suite.

Dans certains pays, l'échantillon pourrait contenir des écoles dites certaines, c'est-à-dire, pour rappel, des écoles dont la mesure de taille est supérieure au pas d'échantillonnage et qui, dès lors, quel que soit l'échantillon, sera toujours reprise. Ces écoles ont été regroupées dans une strate dite *strate d'écoles certaines*. Au sein de cette strate, l'échantillonnage de niveau école a disparu puisque toutes les écoles de la strate font partie de l'échantillon. Au niveau école, il n'y a donc aucune variance d'échantillonnage au sein de cette strate spécifique.

Comme mentionné précédemment, les pseudo-strates doivent être formées au premier niveau auquel un processus d'échantillonnage apparaît. Ainsi, si des provinces, états ou régions sont d'abord sélectionnés et qu'ensuite, au sein des regroupements géographiques retenus, des écoles sont sélectionnées, alors les paires seront constituées au niveau des régions et non au niveau des écoles.

Dans le cas des écoles certaines, le processus d'échantillonnage apparaît pour la première fois au niveau de la classe. Cependant, on ne dispose par école d'une seule classe. Pour estimer adéquatement la variance d'échantillonnage, il aurait fallu échantillonner au moins deux classes dans chacune des écoles certaines.

Pour contourner cette difficulté, il existe deux solutions qui malheureusement conduisent chacune à un léger biais dans l'estimation de la variance d'erreur :

1. on forme les pseudo-strates au niveau des écoles et la variance entre écoles au sein de la strate *écoles certaines* contribuera à la variance d'échantillonnage, qui, par conséquent, sera légèrement estimée ;
2. on scinde les élèves testés de la classe en deux sous-groupes selon une procédure aléatoire et on forme une pseudo-strate au départ de ces deux sous-groupes. Dans ce cas, on ignore la variance entre classes au sein des écoles et on sous-estime la variance d'échantillonnage.

De par la très faible variance attendue entre classes au sein des écoles, dans le cadre de l'étude PASEC, il est conseillé d'implémenter la deuxième solution, à savoir les élèves des écoles certaines sont scindés en deux sous-groupes de taille égale et une pseudo-strate est ainsi constituée par école certaine.

Dans chaque pseudo-strate formée, un indicateur u_i aléatoirement codé 0 ou 1 est attribué à chacune des deux écoles de telle sorte qu'une école soit codée 0 et l'autre 1. Cet indicateur détermine si le poids des élèves échantillonnés dans cette école seront doublés ($u_i=1$) ou seront remis à zéro ($u_i=0$) pour créer la réplique de l'échantillon correspondant à la pseudo-strate.

Au total, 100 réplifications seront produites au primaire par pays. Si un pays, pour des raisons de politiques nationales notamment, sur-échantillonne, les pseudo-strates dont le numéro est supérieur à 100 seront fusionnées. Dans ce cas de figure, 100 réplifications seront produites.

Si un pays se retrouve avec moins de 100 pseudo-strates au primaire et relativement 90 au collège, 100 réplifications (resp. 90 au collège) seront néanmoins construites. Prenons comme exemple un pays qui aurait 80 pseudo-strates. Pour les 80 premières réplifications, le poids des élèves d'une des deux écoles de la pseudo-strate respective seront mis à zéro, le poids des élèves de l'autre école de la pseudo-strate seront doublés. Pour les dix dernières réplifications, le poids des élèves de toutes les écoles participantes sera inchangé. L'estimation du paramètre statistique pour ces dix dernières réplifications sera parfaitement identique à l'estimation obtenue avec les poids finaux. Les différences entre, d'une part, les estimations obtenues par les réplifications avec d'autre part, l'estimation finale, seront toutes égales à 0. La variance d'échantillonnage ne sera donc pas modifiée.

Grâce à cette propriété du JK2, un nombre identique de réplifications peut être construit pour chaque pays, indépendamment du nombre de pseudo-strates. Procéder de la sorte facilite grandement les analyses statistiques.

Pour l'échantillon de 2^e primaire, 50 réplifications seront construites.

Concrètement, les étapes du JK2 sont appliquées directement sur la base de données élève incluant les performances des élèves, les variables de contexte et les pondérations des élèves. À chaque élève j de la base de données est attribué un vecteur de 100 (resp. 90 au collège) poids de réplification (en anglais, *replicate weights*), nommés $wgt_{1j}, wgt_{2j}, wgt_{3j}, \dots, wgt_{100j}$. Ces *replicate weights* sont les transformations du poids final wgt_{0j} de l'élève j . Le poids de la h e réplification est calculé comme suit :

$$wgt_{hj} = wgt_{0j} * k_{hj}$$

où :

$$k_{hj} = \begin{cases} 2 * u_i & \text{si l'élève } j \text{ est dans l'école } i \text{ de la pseudo-strate } h \\ 1 & \text{dans les autres cas} \end{cases}$$

k_{hj} vaudra 2 ($u_i=1$) pour une des deux écoles de la pseudo-strate h et k_{hj} vaudra 0 ($u_i=0$) pour la seconde école. Pour toutes les écoles n'appartenant pas à la pseudo-strate h , la pondération est inchangée. Ainsi, la h e réplification de l'échantillon contient autant d'élèves que l'échantillon principal. Les pondérations des élèves sont inchangées hormis les pondérations des élèves des deux écoles de la pseudo-strate h . La pondération des élèves d'une des deux écoles est doublée, tandis que la pondération des élèves de l'autre école est ramenée à zéro.

La procédure de calcul pour l'estimation d'une statistique, quelle qu'elle soit (moyenne, coefficient de régression, quartile...) se fait comme suit :

- la statistique est calculée pour l'échantillon complet en pondérant les données avec le poids final wgt_{0j} ;
- la statistique est calculée pour chacune des 100 (resp. 90 au collège) répliques d'échantillon (ou 50 pour la 2^e) en utilisant les pondérations respectives ;
- la variance d'échantillonnage est calculée comme suit :

$$\sigma^2_{(\hat{\theta})} = \sum_{i=1}^{90} (\hat{\theta}_{(i)} - \hat{\theta})^2$$

Bibliographie

- Adams, R. & Wu M. (2003). *Programme for International Student Assessment (PISA): PISA 2000 Technical Report*. OECD Publishing, (eds.), Paris. <https://doi.org/10.1787/9789264199521-en>
- Ardilly, P. (2006). *Les techniques de sondage*. Éditions Technip.
- Bertail, P., & Combris, P. (1997). Bootstrap généralisé d'un sondage. *Annales d'économie et de statistique*, 49-83.
- Bloom, B. S. (1979). *Caractéristiques individuelles et apprentissages scolaires* (V. de Landsheere, trad.). Labor ; Nathan. (Ouvrage original publié en 1976).
- Binder, D. A. (1983). On the Variances of Asymptotically Normal Estimators from Complex Surveys. *International Statistical Review / Revue Internationale de Statistique*, vol. 51, no. 3, pp. 279-92. JSTOR. <https://doi.org/10.2307/1402588>
- Cochran, W. G. (1977). *Sampling techniques* (3^e éd.). John Wiley & Sons.
- Davison, A. C., & Sardy, S. (2006). Méthodes de rééchantillonnage pour l'estimation de variance en sondage. *Journal de la Société française de statistique*, 147(3), 3-32.
- Dove, M., & Heck, K. (2020). *Survey Data Analysis Made Easy with SAS* (Paper 4635-2020). SAS Global Forum 2020. <https://support.sas.com/resources/papers/proceedings20/4635-2020.pdf>
- Durbin J. (1967). Design of Multi-Stage Surveys for the Estimation of Sampling Errors. *Journal of the Royal Statistical Society Series C: Applied Statistics*, Volume 16, Issue 2, Pages 152-164. <https://doi.org/10.2307/2985777>
- Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, 7(1), 1-26.
- Gustafsson, J.-E. (2007). Understanding causal influences on educational achievement through analysis of differences over time within countries. Dans : T. Loveless (dir.), *Lessons learned: What international assessments tell us about math achievement* (p. 37-63). Brookings Institution Press.
- Hansen, M. H., & Hurwitz, W. N. (1943). On the theory of sampling from finite populations. *The Annals of Mathematical Statistics*, 14(4), 333-362. JSTOR. <http://www.jstor.org/stable/2235923>
- Horvitz, D. G., & Thompson, D. J. (1952). A Generalization of Sampling Without Replacement from a Finite Universe. *Journal of the American Statistical Association*, 47(260), 663-685. <https://doi.org/10.1080/01621459.1952.10483446>
- IBM Corporation. (2021). *IBM SPSS Complex Samples 28*. IBM Documentation. https://www.ibm.com/docs/en/SSLVMB_28.0.0/pdf/fr/IBM_SPSS_Complex_Samples.pdf
- Johnson, E. G., & Rust, K. F. (1992). Population inferences and variance estimation for NAEP data. *Journal of Educational Statistics*, 17(2), 175-190.
- Kish, L., & Hess, I. (1959). On Variances of Ratios and their Differences in Multi-Stage Samples. *Journal of the American Statistical Association*, 54(286), 416-446. <https://doi.org/10.1080/01621459.1959.10501988>
- Kish, L. (1965). *Survey sampling*. John Wiley & Sons.
- Martin, M. O., Mullis, I. V. S., & Kennedy, A. M. (2003). *PIRLS 2001 Technical Report*. International Study Center, Lynch School of Education, Boston College.
- Neyman, J. (1934). On the Two Different Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection. *Journal of the Royal Statistical Society*, 97(4), 558-625. <https://doi.org/10.2307/2342192>
- OECD. (2009). *PISA data analysis manual: SPSS* (2^e éd.). OECD Publishing.

- Quenouille, M. H. (1949). Problems in plane sampling. *The Annals of Mathematical Statistics*, 20(3), 355-375.
- Rao, J. N. K., & Wu, C. F. J. (1988). Resampling inference with complex survey data. *Journal of the American Statistical Association*, 83(401), 231-241. <https://doi.org/10.1080/01621459.1988.10478591>
- Rust, K. F., & Rao, J. N. K. (1996). Variance estimation for complex surveys using replication techniques. *Statistical Methods in Medical Research*, 5(3), 283-310.
- Särndal, C.-E., Swensson, B., & Wretman, J. (1992). *Model assisted survey sampling*. Springer.
- Srinath, K. P., & Hidiroglo M. A. (1977). *Estimation de la variance dans un sondage à plusieurs degrés*. Techniques d'enquête, 3(1), 1-9. Statistique Canada, n° 12-001-X au catalogue.
- StataCorp. (2019). *Stata survey data reference manual: Release 16*. Stata Press. <https://www.stata.com/manuals16/svy.pdf>
- Tukey, J. W. (1958). Bias and confidence in not-quite large samples. *The Annals of Mathematical Statistics*, 29(2), 614.
- von Davier, M., Mullis, I. V. S., Fishbein, B., & Foy, P. (Eds.) (2023). *Methods and Procedures: PIRLS 2021 Technical Report*. Boston College, TIMSS & PIRLS International Study Center. <https://pirls2021.org/methods>
- Wolter, K. M. (2007). *Introduction to variance estimation* (2^e éd.). Springer.

Les évaluations internationales PASEC : de 2014 à 2024

Évolution du dispositif, spécificités des cycles et principales innovations

Depuis la réforme méthodologique engagée au début des années 2010, le **Programme d'analyse des systèmes éducatifs de la CONFEMEN (PASEC) conduit des évaluations internationales comparatives à grande échelle** afin de produire des données fiables et comparables sur les apprentissages, leur environnement et l'équité des systèmes éducatifs. Les cycles PASEC2014, PASEC2019 et PASEC2024 s'inscrivent dans une même architecture méthodologique, tout en intégrant progressivement de nouvelles populations, de nouveaux champs d'analyse et de nouvelles modalités de collecte. Cette évolution répond à un double objectif : préserver la comparabilité dans le temps et accroître la capacité du dispositif à éclairer le pilotage des politiques éducatives.

1. Un socle méthodologique commun

Les trois cycles reposent sur une évaluation externe et standardisée des apprentissages, conduite selon des procédures harmonisées dans les pays participants. Les populations scolaires sont étudiées à partir d'échantillons probabilistes représentatifs, permettant d'estimer les performances au niveau national sans procéder à une enquête exhaustive.

Le dispositif combine des tests de compétences en langue d'enseignement/lecture et en mathématiques avec des questionnaires contextuels. Les scores sont positionnés sur des échelles internationales et interprétés à travers des niveaux de compétences. Les informations contextuelles recueillies auprès des différents acteurs permettent de relier les performances aux caractéristiques des élèves, des enseignants, des établissements et de leur environnement.

La comparabilité repose également sur des procédures d'assurance qualité couvrant la conception et la mise à l'essai des instruments, leur adaptation linguistique, l'échantillonnage, l'administration, le contrôle de la collecte, le traitement des données et les analyses psychométriques. Les équipes nationales PASEC sont associées à ces différentes étapes, dans une logique de renforcement durable des capacités nationales en évaluation.

2. PASEC2014 : établir la référence internationale

PASEC2014 constitue le premier cycle international standardisé du Programme. Mis en œuvre dans dix pays d'Afrique subsaharienne — Bénin, Burkina Faso, Burundi, Cameroun, Congo, Côte d'Ivoire, Niger, Sénégal, Tchad et Togo — il évalue les élèves en début de scolarité primaire (2^e année) et en fin de scolarité primaire (6^e année) en langue d'enseignement/lecture et en mathématiques. Des données contextuelles sont parallèlement recueillies sur les élèves, les enseignants, les directeurs, les classes et les écoles.

En début de primaire, les épreuves sont administrées individuellement afin d'apprécier précocement les compétences fondamentales. En fin de primaire, l'évaluation repose principalement sur des épreuves papier-crayon et des cahiers tournants. Les résultats sont analysés en termes de performance, de niveaux de compétences, d'efficacité et d'équité des systèmes éducatifs.

L'innovation structurante de PASEC2014 est la création d'une métrique internationale commune. Le cycle transforme les expériences nationales antérieures du PASEC en un dispositif comparatif harmonisé et fournit une première référence commune pour suivre les systèmes éducatifs participants dans l'espace et dans le temps. L'évaluation simultanée du début et de la fin du primaire permet en outre de documenter à la fois les premiers apprentissages et les compétences indispensables à la poursuite de la scolarité.

3. PASEC2019 : consolider les tendances et intégrer les enseignants

PASEC2019 constitue le deuxième cycle international et élargit la participation à quatorze pays. Aux dix pays de 2014 s'ajoutent le Gabon, la Guinée, Madagascar et la République démocratique du Congo. L'architecture

d'évaluation des élèves en début et en fin de primaire est maintenue, ce qui permet, pour les dix pays communs aux deux cycles, d'analyser les évolutions des performances et des disparités entre 2014 et 2019.

Les questionnaires contextuels sont également révisés afin de mieux décrire l'environnement d'apprentissage : questionnaire Élève, questionnaire Enseignant/Classe et questionnaire Directeur/École. Les procédures d'adaptation linguistique, de mise à l'essai et d'assurance qualité sont renforcées afin de sécuriser la comparabilité nationale et internationale des instruments.

La principale innovation de PASEC2019 est l'introduction d'une enquête internationale sur les enseignants du primaire. Celle-ci mesure leurs connaissances disciplinaires en compréhension de l'écrit et en mathématiques ainsi que leurs connaissances didactiques, notamment leur capacité à planifier une situation d'apprentissage et à identifier ou analyser les erreurs des élèves. Elle recueille également des informations sur la formation, l'expérience professionnelle et les conditions de travail. Le diagnostic ne porte donc plus seulement sur les acquis des élèves et leur contexte : il intègre directement les connaissances et compétences professionnelles des enseignants, avec des implications pour les politiques de formation initiale et continue.

4. PASEC2024 : élargir le champ de l'évaluation et moderniser la collecte

PASEC2024 constitue le troisième cycle international et le plus large conduit par le Programme à ce jour, avec seize pays. Il conserve les composantes du primaire nécessaires au suivi des tendances, tout en élargissant le périmètre géographique (16 pays), linguistique et méthodologique de l'évaluation.

La première innovation majeure est l'extension de l'évaluation au collège / à la fin du premier cycle du secondaire. PASEC2024 permet ainsi d'appréhender plus largement le continuum de l'éducation de base, en complétant des mesures réalisées en début et en fin de primaire. Cette extension renforce notamment la contribution du dispositif au suivi de la cible 4.1.1 de l'ODD 4, qui porte sur les niveaux minimaux de compétence à différents moments de la scolarité.

PASEC2024 introduit aussi une innovation linguistique importante. Le cycle marque à la fois une évolution des langues internationales mobilisées, avec l'intégration du portugais et l'élargissement de l'anglais notamment au Nigéria, et une prise en compte accrue des langues nationales utilisées comme langues d'enseignement, particulièrement dans les premières années de scolarité. Cette double évolution permet de mieux refléter la réalité des contextes éducatifs multilingues, de rapprocher les instruments des conditions effectives d'apprentissage et de renforcer l'équité de la mesure, tout en exigeant un travail plus poussé d'harmonisation, d'adaptation et de vérification linguistique afin de préserver l'équivalence des items entre versions.

Le cycle 2024 enrichit également l'analyse de l'environnement des apprentissages. La composante consacrée aux enseignants, introduite en 2019, est maintenue et renforcée ; les chefs d'établissement demeurent une source centrale d'informations sur les établissements ; les parents ou familles sont intégrés parmi les répondants afin de mieux documenter l'environnement familial, le soutien aux apprentissages et certains déterminants extrascolaires. Une attention plus explicite est par ailleurs accordée à l'inclusion et à la prise en compte de la diversité des élèves dans les procédures d'administration.

Une autre innovation structurante concerne la transformation numérique de la collecte. Pour la première fois dans une évaluation internationale PASEC, les élèves de début de primaire sont évalués au moyen de tablettes et d'une application CAPI (Computer-Assisted Personal Interviewing). Cette évolution modernise la chaîne de collecte, facilite certains contrôles de cohérence et contribue à la sécurisation et à la standardisation des données, tout en maintenant des dispositifs adaptés aux différentes composantes de l'enquête.

5. Lecture synthétique de l'évolution des trois cycles

Dimension	PASEC2014	PASEC2019	PASEC2024
Positionnement	1 ^{er} cycle international standardisé	2 ^e cycle ; consolidation et tendances	3 ^e cycle ; changement d'échelle et modernisation
Pays participants	10	14	16
Niveaux élèves	Début et fin du primaire	Début et fin du primaire	Début primaire, fin primaire et collège / fin de l'éducation de base
Domaines élèves	Langue/lecture et mathématiques	Langue/lecture et mathématiques	Lecture/langue et mathématiques, avec extension au collège
Enquête enseignants	Données contextuelles sur les enseignants	Tests disciplinaires et didactiques + questionnaire enseignant	Composante enseignants renforcée ; volume nettement accru
Familles / parents	Informations familiales principalement via les élèves	Informations familiales principalement via les élèves	Parents/familles intégrés comme répondants
Chefs d'établissement	Questionnaires directeurs	Questionnaire Directeur/École	Chefs d'établissement intégrés à la collecte
Collecte	Principalement papier-crayon ; administration individuelle en début de primaire	Papier-crayon ; cahiers tournants ; procédures QA renforcées	Introduction du CAPI/tablettes en début de primaire ; dispositifs papier et numériques selon les composantes
Nombre et types de langues d'évaluation	3 langues : français, anglais et kirundi	7 langues : français, anglais, arabe et langues nationales africaines	14 langues d'enseignement, combinant langues internationales/officielles, dont l'anglais et le portugais, et langues nationales
Langues nationales prises en compte	Kirundi	Kirundi, malagasy, haoussa et zarma	Élargissement à plusieurs langues nationales, dont kirundi, malagasy, haoussa, zarma, bamanankan, wolof, soninké, sérère, pulaar et mandingue
Innovation et adaptation linguistique	Première prise en compte de contextes linguistiques différenciés dans un cadre international commun	Diversification des versions et formalisation des procédures de traduction, conciliation, vérification et validation	Changement d'échelle multilingue : extension des langues internationales, meilleure prise en compte des langues nationales d'enseignement et renforcement de l'assurance qualité linguistique
Comparabilité temporelle	Création de la référence	Mesure des tendances 2014–2019 pour les pays communs	Continuité des séries du primaire + nouvelles composantes
Innovation dominante	Métrique internationale commune	Évaluation des compétences des enseignants	Collège, familles, numérique, inclusion, élargissement géographique et linguistique
Lien avec l'ODD 4	Contribution à la mesure des apprentissages	Suivi renforcé des apprentissages et de l'équité	Couverture élargie de la cible 4.1.1 jusqu'à la fin du premier cycle du secondaire

En synthèse, PASEC2014 a standardisé et créé la référence internationale ; PASEC2019 a consolidé cette référence, rendu possible l'analyse des tendances et intégré les compétences des enseignants ; PASEC2024 conserve les séries comparables tout en élargissant l'évaluation à de nouvelles populations et à la fin de l'éducation de base, et en modernisant

les modalités de collecte. Cette progression traduit une évolution d'un dispositif centré sur la mesure des acquis des élèves vers un système plus complet sur les conditions de production des apprentissages et les leviers d'amélioration de la qualité de l'éducation.

PASEC

Évaluer les apprentissages pour transformer les politiques éducatives

Créé en 1991, le Programme d'analyse des systèmes éducatifs de la CONFEMEN (PASEC) s'est imposé comme l'instrument de référence au service du pilotage des systèmes éducatifs des États et gouvernements membres de la CONFEMEN. Il a pour vocation d'éclairer la décision publique en mesurant l'évolution des performances des systèmes éducatifs, pour orienter l'élaboration, le suivi et l'évaluation des politiques éducatives et, à travers elles, l'amélioration de la qualité de l'éducation.

Son action ne se limite pas à la mesure. Le PASEC renforce les capacités nationales d'évaluation, favorise l'appropriation et l'usage effectif des résultats par les acteurs éducatifs, et soutient l'essor de la recherche en éducation. Par ses évaluations et ses travaux d'analyse, il met à la disposition des décideurs des données fiables et comparables ainsi que des analyses rigoureuses, leviers d'une éducation plus performante, plus équitable et plus efficace.

Le PASEC concourt enfin à la production de connaissances en accompagnant les chercheurs et en mettant à leur disposition ses données et ses cadres d'évaluation. Il nourrit des travaux utiles tant aux systèmes éducatifs qu'à la communauté universitaire.

CONFEMEN – PASEC

Complexe SICAP Point E, Immeuble C
3^e étage, avenue Cheikh Anta Diop
BP 3220 Dakar, Sénégal
Tél. : +221 33 859 29 79 / +221 33 859 29 91
Fax : +221 33 825 17 70
Courriel : pasec@confemen.org
Site web : www.pasec.confemen.org

@CONFEMEN



Avec le soutien de

