

RAPPORT TECHNIQUE DE L'ÉVALUATION INTERNATIONALE PASEC2019



pasec
Programme d'analyse des systèmes
éducatifs de la confemen

Conférence des ministres de l'Éducation des États et Gouvernements de la Francophonie

Merci de citer cette publication comme suit :

PASEC (2022). *Rapport technique de l'évaluation internationale PASEC2019*. PASEC, CONFEMEN, Dakar.

**RAPPORT TECHNIQUE
DE L'ÉVALUATION
INTERNATIONALE
PASEC2019**

Sigles et acronymes

ACP	Analyse en composantes principales
CCI	Courbe caractéristique de l'item
CONFEMEN	Conférence des ministres de l'Éducation des États et gouvernements de la Francophonie
DIF	Fonctionnement différentiel des items (<i>Differential Item Functioning</i>)
ECIM-E	Épreuve collective d'identification des mots écrits
IRT	Théorie de réponse à l'item (<i>Item Response Theory</i>)
MLMCM	Modèle logit multinomial à coefficients mixtes
MNSQ	Carré moyen (<i>Mean Square</i>)
MOS	Mesure de taille (<i>Measure of Size</i>)
NA	Nombre aléatoire
OCDE	Organisation de Coopération et de Développement Économiques
PASEC	Programme d'Analyse des Systèmes Éducatifs de la CONFEMEN
PIB	Produit intérieur brut
PISA	Programme International pour le Suivi des Acquis des élèves
QCM	Question à choix multiples
SES	Statut socioéconomique (<i>Socioeconomic Status</i>)
TIMSS	<i>Trends in International Mathematics and Science Study</i>
ULMCM	Modèle unidimensionnel logit multinomial à coefficients mixtes
VP	Valeurs plausibles
WLE	<i>Weighted Likelihood Estimation</i>

Table des matières

1. TESTS COGNITIFS	10
Fondements des tests PASEC2019 de début et fin de scolarité primaire ainsi que de l'enquête sur les enseignants du primaire	12
Développement des tests PASEC2019	13
Partie I. Tests PASEC2019 de début de scolarité primaire	14
1.1.1 Test de langue de début de scolarité primaire	14
1.1.2 Test de mathématiques de début de scolarité primaire	15
1.1.3 Protocole et spécificités des tests PASEC2019 de début de scolarité primaire	16
Partie II : Tests PASEC2019 de fin de scolarité primaire	17
1.2.1 Test de lecture de fin de scolarité primaire	17
La compréhension de mots et de phrases isolés et la compréhension de textes	18
1.2.2 Test de mathématiques de fin de scolarité primaire	21
1.2.3 Protocole et spécificités des tests PASEC2019 de fin de scolarité primaire	25
Partie III. Test des enseignants du primaire dans le cadre de l'évaluation PASEC2019	26
1.3.1 Les domaines disciplinaires et didactiques en compréhension de l'écrit et en mathématiques et leurs compétences évaluées du test PASEC2019 des enseignants primaires	26
1.3.2 Processus cognitifs de contenu et didactique de compréhension de l'écrit et de mathématiques dans l'évaluation PASEC2019 sur les enseignants	27
1.3.3 Protocole et spécificités des tests PASEC2019 des enseignants du primaire	28
2. QUESTIONNAIRES CONTEXTUELS	30
5.1. Approche globale du PASEC pour étudier le contexte éducatif	34
5.2. Cadre théorique des questionnaires de contexte	36
5.3. Spécifications sur les questionnaires de contexte	40
3. PLAN D'ÉCHANTILLONNAGE	42
3.1 Procédures d'échantillonnage du PASEC2019	44
3.1.1 Couverture de la population cible et taux de participation au niveau des écoles et des élèves	45
3.1.2 Taux de participation au niveau des écoles	46
3.1.3 Taux de participation au niveau des élèves	47
3.2 Procédures techniques d'échantillonnage des écoles pour les pays	48
3.2.1 Définition de la population cible nationale	48
3.2.2 Élaboration de la base de sondage	48
3.2.3 Définition des variables de stratification	49
3.2.4 Attribution d'une mesure de taille à chaque école	51
3.2.5 Calcul des taux d'exclusion d'écoles	51
3.2.6 Traitement des petites écoles et des très petites écoles	51
3.2.7 Échantillonnage des écoles	52
3.3 L'échantillonnage des classes et des élèves dans les pays participants	56
3.4 Attribution des cahiers aux élèves de la 6^e année et aux enseignants des écoles	56
4. COLLECTE ET SAISIE DES DONNÉES DANS LES PAYS	57
4.1 Collecte des données	58
4.1.1 Préparation de la collecte des données	58
4.1.2 Formation des administrateurs de tests	60
4.1.3 Collecte des données	62
4.2 Rangement et sécurisation des instruments de collecte revenus du terrain	66
4.3 Saisie des données	68
4.3.1 Codification des instruments des tests de 6 ^e année	68
4.3.2 Codification des instruments des tests de 2 ^e année	68
4.3.3 Codifications des instruments des questionnaires contextuels	69
4.3.4 Développement des masques de saisie	69
4.3.5 Recrutement et formation des agents de codification et de saisie	70
4.3.6 Saisie des données et contrôle de qualité	70
4.3.7 Sauvegarde des données	72

4.3.8	Communication entre les équipes nationales et le PASEC	72
5.	NETTOYAGE DES DONNÉES	74
5.1.	Programmation sous Stata des procédures de nettoyage des données et identification des incohérences dans les données	77
5.2.	Recodage des données avec l'appui des équipes nationales	78
5.2.1	Vérification et correction des incohérences au niveau des identifiants dans chaque base de données	78
5.2.2	Vérification et correction des incohérences entre les différentes bases de données	78
5.2.3	Vérification et correction des incohérences au niveau interne des bases de données	79
5.3.	Mission de finalisation du nettoyage des données dans les pays	80
5.4.	Finalisation des bases de données : Recodage des items (non atteints et manquants par design)	80
6.	PONDÉRATION DES DONNÉES ET CALCUL DES POIDS ISSUS DE LA RÉPLICATION	81
6.1	Pondération des données et calcul des taux de participation	83
6.1.1	Pondération des données	83
6.1.2	Le poids initial de l'école	84
6.1.3	L'ajustement pondéral pour la non-réponse des écoles	85
6.1.4	Le poids de la classe au sein de l'école	85
6.1.5	Le poids initial de l'élève au sein de sa classe	85
6.1.6	L'ajustement pour la non-réponse des élèves	86
6.1.7	Calcul des taux de participation des écoles et des élèves	86
6.1.8	Taux de participation des écoles avant remplacement	86
6.1.9	Taux de participation des écoles après remplacement	87
6.2	Calcul de la variance d'échantillonnage	90
6.2.1	La construction des zones de Jackknife, ou pseudo-strates	91
6.2.2	La génération des poids de réplifications	91
7.	MISE À L'ÉCHELLE DES DONNÉES COGNITIVES	94
7.1	Scaling model	95
7.2	Item analysis and calibrations	95
7.3	Item response model fit (weighted mean square MNSQ)	95
7.4	Discrimination coefficients	96
7.5	National reports	96
7.6	Item-by-country interaction	100
7.7	International calibration	104
7.8	Test targeting	107
7.9	Linking of PASEC 2019 to PASEC 2014 scale for Grade 2 and Grade 6	115
7.10	Scaling models and Reliability	117
7.11	Correlations between Main Domains	118
7.12	Student score generation	119
7.13	Constructing conditioning variables	120
8.	ÉLABORATION DES ÉCHELLES DE COMPÉTENCES DES ÉLÈVES	157
8.1.	Méthodologie adoptée	158
8.2.	Niveaux de compétence et points de césure	162
4.3.2	Transformation linéaire pour passer de l'échelle <i>Logit</i> à l'échelle 500/100	164
8.3.	Description des échelles et des niveaux de compétences	164
8.4.	Procédure de création des indices contextuels	169
	Recodage des données brutes	170
	Cohérence interne des indices	171
	Analyse psychométrique des paramètres d'item des indices par pays	171
	Synthèse des analyses nationales	173
	Estimation internationale des paramètres d'item	173
	Estimation des scores	174
10.	CALCUL DE L'ERREUR TYPE	177
10.1	Sources d'erreur : Erreur d'échantillonnage et erreur de mesure	179
	Erreur d'échantillonnage	179
	Erreur de mesure	179
	Erreur totale	180

10.2 Nature des échantillons et calcul de l'erreur type d'une différence	181
Échantillons indépendants	181
Échantillons dépendants	181
10.3 Quelques exemples de calcul de statistiques	183

1. TESTS COGNITIFS

Les épreuves du PASEC s'intéressent aux savoirs et savoir-faire indispensables aux élèves pour poursuivre avec succès leurs apprentissages dans l'enseignement primaire et secondaire ou professionnel, mais aussi aux capacités des élèves à mobiliser leurs compétences en lecture et en mathématiques pour comprendre, apprendre et s'intégrer au quotidien.

Ce deuxième cycle de l'évaluation internationale PASEC2019 propose un cadre méthodologique en phase avec les standards scientifiques internationaux en matière d'évaluation en langue-lecture et en mathématiques et adapté aux contextes et enjeux spécifiques, en début et en fin de scolarité primaire, des systèmes éducatifs des pays PASEC de la CONFEMEN. Aussi, le PASEC assure les mêmes exigences méthodiques autour de l'enquête sur les enseignants du primaire introduite dans cette nouvelle édition de l'évaluation. Les cadres de référence des tests PASEC de début et de fin de scolarité primaire et de l'enquête sur les enseignants visent à documenter les orientations méthodologiques des différentes épreuves PASEC. Ce chapitre présente de manière synthétique ces documents.

Le lecteur est invité à consulter ces documents sur le site www.pasec.confemen.org pour plus d'information sur les fondements, les contenus et les spécificités des différents tests.

Fondements des tests PASEC2019 de début et fin de scolarité primaire ainsi que de l'enquête sur les enseignants du primaire

Les tests PASEC sont construits sur la base :

- i. de recherches scientifiques dégageant les différents stades d'apprentissage de la lecture et des mathématiques ;
- ii. des niveaux de compétence en lecture et en mathématiques des élèves, du contexte des pays évalués et des principaux domaines d'enseignement en vigueur dans les programmes scolaires des pays ;
- iii. des standards de mesure¹ en lecture et en mathématiques couramment utilisés au niveau international.

Au regard de ces spécificités, les tests PASEC n'évaluent pas spécifiquement le degré de maîtrise des curricula nationaux par les élèves mais contribuent à évaluer les capacités des élèves à atteindre des objectifs plus généraux (les « compétences clés ») basés sur un référentiel en langue-lecture et en mathématiques commun, universel et adapté aux enjeux de l'école et des sociétés modernes.

Dès lors, l'évaluation PASEC doit être considérée comme une mesure externe et complémentaire aux évaluations nationales qui fixent leurs propres standards en fonction des objectifs spécifiques des systèmes éducatifs nationaux.

Quant aux tests relatifs aux enseignants ils évaluent, pour une part, les compétences attendues en fin de primaire, quelle que soit la classe de primaire encadrée par l'enseignant. Cette option repose sur le principe qu'un enseignant de primaire doit être capable d'enseigner à tous les niveaux du cycle car il peut être à tout moment affecté dans un niveau différent que celui qu'il tient. Elle repose aussi sur l'exigence que tout enseignant doit avoir en tête le profil d'un élève de fin de 6^{ème} année du point de vue des compétences à atteindre.

¹ Les standards internationaux de mesure font référence aux procédures de construction, d'administration et d'analyse des tests.

D'autre part, les tests enseignants évaluent une dimension spécifique du métier à savoir l'utilisation des savoirs didactiques à travers deux compétences professionnelles majeures : planifier une situation d'apprentissage et identifier les types d'erreur des élèves.

Développement des tests PASEC2019

Les items des tests ont été conçus en français² par le centre international du PASEC et ont été validés par son Comité scientifique. Un comité d'experts du Centre de recherche en éducation de l'Université de Liège et de Genève, de concert avec les équipes nationales PASEC des 14 pays participant à l'évaluation internationale PASEC2019, ont contribué à la mise en place de ces instruments de mesure. Le développement des tests a suivi un processus scientifique conforme aux standards des évaluations internationales.

La traduction et l'adaptation³ des tests dans d'autres langues que le français⁴ ont été réalisées par un groupe de traducteurs nationaux et experts⁵ dans l'adaptation linguistique d'items dans les évaluations internationales. Le processus d'adaptation implique une traduction puis une vérification indépendante.

Deux laboratoires cognitifs distincts ont été menés au Togo autour des instruments de test des élèves et au Niger autour de ceux de l'enquête sur les enseignants dans le but d'observer, prétester et ajuster lesdits instruments. Les procédures de passation et le fonctionnement des items ont été expérimentés en avril 2018 dans tous les pays participants auprès d'un échantillon réduit de 20 écoles. Cette phase a permis d'apprécier le fonctionnement global des tests et le fonctionnement individuel des items aux niveaux national et international. Les items défectueux ont été ajustés ou supprimés en fonction de leurs caractéristiques psychométriques (voir le chapitre 7 pour plus d'information sur ces aspects).

Les tests définitifs ont été validés en novembre 2018 puis administrés dans tous les pays en avril et mai 2019 dans un échantillon d'écoles. Comme lors de la phase d'expérimentation, les items ont été analysés pour apprécier le fonctionnement global des tests et le fonctionnement individuel des items aux niveaux national et international. Les items retenus après les analyses psychométriques ont été intégrés aux échelles de scores internationales PASEC2019 (voir le chapitre 7 pour plus d'information sur ces aspects).

Au terme de l'évaluation PASEC2019, un échantillon d'items a été rendu public dans le rapport international pour illustrer les résultats des élèves. Un échantillon d'items de l'évaluation PASEC2019 est préservé et permettra d'établir une comparaison dans le temps des performances des élèves lors des prochains cycles d'évaluation du PASEC. Un nouveau lot d'items sera développé pour la prochaine évaluation internationale pour combler les items libérés.

² Langue source de conception des items.

³ Les tests ne sont pas traduits directement mais adaptés, de telle sorte que la nature de la question s'approche autant que possible de celle de la version source en français.

⁴ Pour le PASEC2019, le test a été adapté en anglais pour le sous-système anglophone camerounais.

⁵ La société CAPSTAN a été mandatée pour réaliser l'adaptation des tests PASEC2019 en anglais.

Partie I. Tests PASEC2019 de début de scolarité primaire

Les tests PASEC2019 de début de scolarité primaire sont administrés individuellement aux élèves de 2^e année du primaire pour mesurer les compétences acquises au cours des premiers apprentissages de la langue d'enseignement et des mathématiques et faire au plus tôt le diagnostic du premier bilan de leurs compétences fondamentales. Le test permet également d'identifier les difficultés d'apprentissage avant qu'elles n'entraînent des échecs et des abandons scolaires.

La durée globale des tests est d'environ 30 minutes par discipline.

1.1.1 Test de langue de début de scolarité primaire

En fin de 2^e année du primaire qui correspond pour la plupart des pays à la fin d'un premier sous cycle d'apprentissage, les curricula indiquent que tous les apprentis lecteurs des pays de la CONFEMEN devraient être capables de lire et de comprendre un message court, simple et familier. Cette aptitude implique pour l'élève d'avoir atteint un premier niveau de déchiffrage pour reconnaître les mots familiers, automatiser les mécanismes liés au déchiffrage de l'écrit et posséder un niveau de compréhension orale suffisant dans la langue d'enseignement. Ces différentes capacités sont évaluées dans le nouveau test.

Encadré 1.1 : domaines évalués par le test PASEC2019 de langue – Début de scolarité

L'encadré ci-dessous définit les trois domaines évalués en langue d'enseignement dans le test PASEC de début de primaire. Chacun des trois domaines est évalué successivement dans le test.

Compréhension de l'oral : La compréhension de l'oral est évaluée à travers des messages oraux associant des mots et phrases isolés et des textes. Le développement des compétences dans ce domaine permet aux élèves d'étendre leur vocabulaire pour automatiser le décodage en lecture à travers les correspondances entre l'oral et l'écrit.

Lecture-décodage : La lecture-décodage est évaluée à travers des situations de reconnaissance des caractéristiques de l'écrit. La lecture-décodage est évaluée à travers des situations d'identification graphophonologique (lettre, syllabe et mots) et d'activités aisées de lecture de lettres et de mots. Le développement des compétences dans ce domaine permet aux élèves d'automatiser leur lecture pour déterminer le sens des mots et des phrases, et ainsi étendre leur vocabulaire.

Compréhension de l'écrit : La compréhension de l'écrit est évaluée à travers des situations de lecture de mots et phrases isolés et de textes dans lesquels l'élève est amené à retrouver, combiner et interpréter des informations. Le développement des compétences dans ce domaine permet aux élèves de lire en autonomie dans des situations quotidiennes variées, pour développer leurs savoirs et participer à la vie en société.

Tableau 1 : poids des domaines évalués et nombre d'items dans le test PASEC2019 de langue – Début de scolarité

Domaines	Total général
Compréhension de l'oral	37 % (16)
Décodage	28 % (12)
Compréhension de l'écrit	35 % (15)
Total général	100 % (43)

70 % des items de l'épreuve de langue PASEC2019 (30 des 43 items) sont des items d'ancrage avec le test PASEC2014.

Les compétences liées à la production orale (expression orale, récitation...), à l'écriture (maîtrise de la graphie, production de phrases par écrit...) et à l'étude de la langue (grammaire, orthographe...) ne seront pas évaluées par le test PASEC de début de primaire en raison des spécificités linguistiques propres à chaque langue. Et des questions techniques notamment liées à la standardisation de l'épreuve d'une langue à l'autre et à la volonté de disposer d'une mesure comparable d'un pays à l'autre quelle que soit la langue.

1.1.2 Test de mathématiques de début de scolarité primaire

Les tests PASEC reflètent les processus indispensables pour l'acquisition des compétences fondamentales en arithmétique, en géométrie, espace et mesure, qui permettent à l'élève de passer d'une connaissance analogique et intuitive à une connaissance symbolique des concepts mathématiques⁶. Le but est d'identifier les étapes, lors de l'acquisition des compétences de base, où les élèves éprouvent des difficultés afin d'orienter les politiques éducatives dans la mise en place de remédiation ou dans la réorientation des curricula et des pratiques pédagogiques.

Le choix des compétences évaluées dans le test PASEC de mathématique s'appuie à la fois sur les recherches scientifiques dans le domaine des premiers apprentissages, mais également sur les curricula des pays participant aux évaluations PASEC. Cette revue de la littérature a montré que l'acquisition de certaines compétences en arithmétique était fondamentale. Ainsi, le test PASEC2019 accorde une plus grande pondération à l'arithmétique qu'à la géométrie et au repérage dans l'espace.

⁶ Les compétences évaluées par le PASEC suivent en grande partie les recommandations du *Learning Metrics Task Force* (2013), excepté celle concernant la capacité des élèves à résoudre des « problèmes de configuration » (*pattern problem*), qui relève de la logique. Le PASEC n'évalue pas cette compétence dans le test de mathématiques car les mises à l'essai des items ont été peu probantes quant aux items de logique. Ceci pose d'une part la question de l'enseignement de la logique dans les pays évalués par le PASEC, et d'autre part celle de la façon dont il faut adapter ce type d'items aux contextes nationaux.

Tableau 2 : Domaines et compétences évalués en mathématiques et nombres d'items dans le test de début de scolarité de l'évaluation PASEC2019

Domaines	Sous domaine	Compétences évaluées et exercices	Poids des domaines dans le test
Arithmétique	Acquisition de la chaîne verbale des nombres, le dénombrement et la comparaison des quantités.	<ul style="list-style-type: none"> Compter oralement Identifier des nombres écrits Dénombrer des collections d'objets Discriminer des quantités et classer les nombres Trouver un nombre manquant dans une suite logique de nombres 	72,5 (29)
	Acquisition des outils arithmétiques : l'addition et la soustraction.	<ul style="list-style-type: none"> Résoudre des additions et des soustractions simples et complexes (en dessous de 20 et au-dessus de 20) 	
	Résolution de problèmes	<ul style="list-style-type: none"> Comprendre différents types de problèmes et leurs énoncés. Émettre un raisonnement et d'effectuer une opération simple pour résoudre un problème 	
Géométrie, espace mesure	Connaissance des formes géométriques simples, le repérage dans l'espace, le classement des objets selon leur taille.	<ul style="list-style-type: none"> Connaître les différentes formes géométriques de base (carré, triangle, rectangle) Reconnaître la position des objets dans l'espace (en dessous de, à côté de, dans, sur, etc.) 	27,5 % (11)
Total	-	-	100 % (40)

47,5 % des items de l'épreuve de mathématiques PASEC2019 (19 des 40 items) sont des items d'ancrage avec le test PASEC2014.

1.1.3 Protocole et spécificités des tests PASEC2019 de début de scolarité primaire

Administration des tests

Les tests sont administrés individuellement aux élèves par un administrateur. La passation des tests s'effectue sur quatre matinées et touche au maximum 16 élèves répartis en deux sous-groupes d'une même classe pour chacune des écoles sélectionnées dans le plan d'échantillonnage.

Pour chaque école, l'ordre de passation des tests selon les disciplines, langue ou mathématiques, est défini aléatoirement par le centre international du PASEC en fonction du plan d'échantillonnage et permet une répartition équilibrée de l'ordre de passation des disciplines entre les écoles.

Afin d'alléger la passation de test à l'autre, l'échantillon a été scindé en deux sous- groupes. Les sous- groupes passent le test de langue les premières matinées et enchainent avec ceux de mathématiques et questionnaire les dernières matinées.

Le test de langue s'organise en trois phases successives, chaque phase correspondant à l'évaluation d'un des trois domaines de langue. Chacune des trois parties contient une série d'exercices, et chaque exercice comprend un exemple puis une suite d'items.

Partie II : Tests PASEC2019 de fin de scolarité primaire

Les épreuves PASEC de fin de scolarité primaire s'intéressent aux savoirs et savoir-faire en lecture et en mathématiques indispensables aux élèves pour poursuivre une éducation de qualité dans l'enseignement secondaire ou professionnel, mais aussi aux capacités des élèves à mobiliser leurs compétences pour comprendre, apprendre et s'intégrer dans des situations de leur environnement quotidien.

1.2.1 Test de lecture de fin de scolarité primaire

En fin de cycle, le test PASEC2019 mesure principalement les compétences de compréhension de textes informatifs et de documents. La capacité d'extraire des informations de textes littéraires est aussi évaluée, dans une moindre mesure, et les activités de décodage de mots et de phrases isolés occupent à ce stade une place mineure (voir l'encadré 1.2). Le test ne mesure pas les compétences en production écrite et en compréhension ou expression orale ni les outils (orthographe, grammaire, conjugaison, etc.) propres à chaque langue.

Encadré 1.2 : définition des domaines de compréhension de l'écrit évalués par le PASEC2019

Compréhension de mots et de phrases isolés : La compréhension de mots et de phrases isolés est évaluée à travers des situations de lecture portant sur la découverte du sens explicite de mots et de phrases isolés. Le développement des compétences dans ce domaine permet aux élèves d'automatiser leur lecture pour accéder progressivement au sens des phrases et des textes et pour étendre leur vocabulaire. Le niveau de ces tâches est très basique et est équivalent aux objectifs des programmes scolaires de début de scolarité primaire.

Compréhension de textes : La compréhension de texte est évaluée à travers des situations de lecture de textes narratifs, informatifs et de documents, desquels les élèves sont amenés à extraire, réaliser des inférences simples, interpréter et combiner des informations. Le développement des compétences dans ce domaine permet aux élèves de lire de façon autonome dans des situations quotidiennes variées pour développer leurs savoirs et participer à la vie en société.

Le test est composé de 98 items repartis dans le tableau ci-dessous. Ce test accorde une place centrale à l'évaluation des compétences de lecture de base sur des textes informatifs et des

documents. Les activités de compréhension de mots et de phrases isolés ainsi que la lecture de textes littéraires occupent une place mineure dans le test.

Tableau 3 : poids attendu des domaines de compréhension de l'écrit dans le test PASEC2019

Sous-domaines de lecture	Supports de lecture	Proportion dans le test
Compréhension de mots et de phrases isolés	Images, mots et phrases isolés	16 %
Compréhension de textes	Textes littéraires	39%
	Textes informatifs et documents	45 %

1.2.1.1 Présentation des sous-domaines de contenus évalués dans le test PASEC2019 de lecture de fin de scolarité primaire

La compréhension de mots et de phrases isolés et la compréhension de textes

Sous-domaine de compréhension de mots et de phrases isolés

L'évaluation du niveau de compréhension de mots et de phrases isolés permet d'apprécier la part des élèves qui en fin de scolarité primaire éprouvent de la difficulté avec des tâches de lecture de bas niveau. Ces apprentis lecteurs, qui concentrent leur attention sur l'identification de mots isolés, accèdent difficilement à la compréhension du texte.

Les items visent à mesurer les capacités de compréhension de mots et/ou phrases de la vie quotidienne. L'information à retrouver est explicite et concentrée dans une série de mots isolés ou dans une phrase courte et simple de 10 mots maximum (sujet/verbe/complément).

Sous-domaine de compréhension de textes

Passé le stade de la compréhension de mots et phrases isolés, l'élève mobilise avec plus d'aisance ses capacités pour passer à la compréhension de plusieurs informations, de phrases et de textes afin d'identifier des informations et construire du sens.

Dès lors, le lecteur accède au sens d'un message écrit en mobilisant plusieurs habiletés. Il peut, prélever une information explicite, associer plusieurs informations et les mettre en rapport avec ses connaissances extérieures, synthétiser et analyser un texte, donner son opinion et argumenter sur un texte.

Le cadre de référence PASEC catégorise une partie de ces différentes habiletés dans trois grands groupes nommés processus cognitifs : extraire une information explicite, réaliser des inférences simples, interpréter et combiner des informations.

Comme pour PASEC2014, le test PASEC2019 administré aux élèves comprend à la suite de chaque texte, une série de questions en rapport avec une des trois catégories de processus cognitifs. Chacune des questions est classée exclusivement dans une des trois catégories.

Le tableau suivant présente la répartition des questions dans les épreuves définitives PASEC2019 selon les trois processus cognitifs et les deux catégories de compréhension de texte.

Tableau 5 : poids des processus cognitifs évalués et des supports de lecture en compréhension de texte dans l'évaluation PASEC2019

Poids des processus cognitifs évalués dans PASEC2019	Poids des supports de lecture en compréhension de texte dans PASEC2019		Total compréhension de texte
	Textes littéraires	Textes informatifs et documents	
• Extraire une information explicite	27 % (22)	37 % (30)	(52)
• Réaliser des inférences simples	12 % (10)	8,5 % (7)	(17)
• Interpréter et combiner des informations	7 % (6)	8,5 % (7)	(13)
Total	46 % (38)	54 % (44)	(82)

Les caractéristiques principales des trois processus cognitifs sont décrites et illustrées ci-dessous.

1.2.1.2 Présentation des processus cognitifs évalués dans le test PASEC2019 de lecture de fin de scolarité primaire

Tableau 4 : descriptif des tâches à réaliser dans le processus « décoder et reconnaître une information » du test de lecture de fin de scolarité

Processus cognitif	Tâches à réaliser
Décoder et reconnaître une information	<ul style="list-style-type: none"> • Associer un mot ou une phrase avec une image • Trouver le sens d'un mot ou d'une phrase • Catégoriser du vocabulaire familier

- **Extraire une information explicite**

Les lecteurs extraient des informations pour confirmer ou rejeter des hypothèses qui leur permettent de comprendre ce qu'ils lisent. Dans le cadre du test de compréhension de l'écrit, les élèves sont amenés à extraire une information concrète localisée dans une seule phrase afin de répondre à une question explicitement posée.

Tableau 6 : Descriptif des tâches à réaliser dans le processus cognitif « extraire une information » du test de lecture de fin de scolarité – PASEC2019

Processus cognitif	Tâches à réaliser
Extraire une information explicite	<ul style="list-style-type: none"> • Localiser une information concrète dans une phrase • Identifier les caractéristiques principales d'une histoire (thème, personnage, lieu...) • Reconnaître une information littéralement équivalente à une information explicite

- **Réaliser des inférences simples**

Au fur et à mesure qu'un élève construit du sens par rapport à un texte, il est amené à nuancer davantage son analyse pour réaliser des inférences directes et en déduire des informations complémentaires en se basant sur le sens littéral d'une phrase ou d'un paragraphe. Le lecteur perçoit des éléments concrets qui lui permettent de construire du sens par déduction et raisonnement logique. Il met en relation plusieurs phrases et idées fournies par le texte pour déduire le sens d'une expression, d'une phrase ou, à un niveau plus poussé, l'idée générale d'un passage dans un texte.

Tableau 7 : Descriptif des tâches à réaliser dans le processus « réaliser des inférences simples » du test de lecture de fin de scolarité – PASEC2019

Processus cognitif	Tâches à réaliser
Réaliser des inférences logiques	<ul style="list-style-type: none"> • Identifier la référence d'un pronom ou d'une expression (inférence anaphorique)

-
- Chercher les causes et les conséquences d'un événement (inférence causale, déductive)
 - Repérer l'idée générale explicite d'une phrase ou d'un passage (inférence logique)
 - Mettre en relation plusieurs informations concrètes localisées dans plusieurs phrases pour repérer les caractéristiques d'un texte
-
- **Interpréter et combiner des informations**

Lorsqu'un élève est à l'aise pour réaliser des inférences directes basées sur des informations essentiellement explicites à l'échelle d'une phrase ou d'un paragraphe de texte, il peut être en mesure d'interpréter des idées notamment à l'échelle globale d'un texte, en reliant plusieurs indices situés tout au long du texte tout en s'appuyant sur ses connaissances et expériences antérieures. Les informations données dans le texte nécessitent une mise en relation avec des connaissances personnelles du monde pour créer une nouvelle information.

Ainsi, les inférences que doivent réaliser le lecteur sont plus complexes et plus sous entendues par rapport à celles définies dans le processus « réaliser des inférences simples ». La difficulté repose tout particulièrement sur la capacité du lecteur à mobiliser des connaissances hors du texte et à traiter le texte dans sa globalité.

Tableau 8 : Descriptif des tâches à réaliser dans le processus « interpréter et combiner des informations » du test de lecture de fin de scolarité – PASEC2019

Processus cognitif	Tâches à réaliser
Interpréter et combiner des informations	<ul style="list-style-type: none"> • Interpréter le sens d'une expression impliquant un vocabulaire soutenu (inférence pragmatique) • Interpréter le thème global d'un texte (inférence pragmatique) • Associer plusieurs informations implicites situées dans différentes parties d'un texte pour déduire une nouvelle information • Synthétiser plusieurs informations et les associer avec ses connaissances générales sur la situation • Identifier l'intention de l'auteur ou les sentiments d'un personnage • Sélectionner un titre pour le texte

1.2.2 Test de mathématiques de fin de scolarité primaire

Les tests PASEC reflètent les processus indispensables pour l'acquisition des compétences fondamentales en arithmétique, en géométrie, espace et mesure, qui permettent à l'élève de passer d'une connaissance analogique et intuitive à une connaissance symbolique des concepts mathématiques⁷. Le but est d'identifier les étapes, lors de l'acquisition des compétences de base, où

⁷ Les compétences évaluées par le PASEC suivent en grande partie les recommandations du *Learning Metrics Task Force* (2013), excepté celle concernant la capacité des élèves à résoudre des « problèmes de configuration » (*pattern problem*), qui relève de la logique.

les élèves éprouvent des difficultés afin d'orienter les politiques éducatives dans la mise en place de remédiation ou dans la réorientation des curricula et des pratiques pédagogiques. Le choix des compétences évaluées dans le test PASEC de mathématique s'appuie à la fois sur les recherches scientifiques dans le domaine des premiers apprentissages, mais également sur les curricula des pays participant aux évaluations PASEC. Cette revue de la littérature a montré que l'acquisition de certaines compétences en arithmétique était fondamentale. Ainsi, le test PASEC2019 accorde une plus grande pondération à l'arithmétique qu'à la géométrie et au repérage dans l'espace.

Encadré 1.3 : Sous-domaines évalués par le test PASEC2019 de mathématiques – Fin de scolarité

Nombres et opérations

- Nombres entiers positifs (*ou nombres naturels*) : compréhension des nombres et du système en base 10 (U, D, C, M) ; connaissance et compréhension des priorités des opérations et des propriétés des quatre opérations ; opérations sur les nombres telles additionner, soustraire, multiplier, diviser.
- Nombres décimaux et fractions : compréhension des nombres décimaux et des fractions-nombres ; compréhension de la fraction-opérateur ; compréhension des pourcentages et de leur relation avec les fractions-opérateurs ; opérations sur les nombres décimaux et sur les fractions et opérations impliquant le calcul d'un pourcentage.

Grandeurs et mesures

- Unités de mesures : connaissance et compréhension des unités de mesure conventionnelles telles que les longueurs, les masses, les capacités, les angles, les heures et les durées ; conversion d'unités de mesures.
- Calcul des grandeurs : opérations impliquant des unités de mesures telles que les longueurs, les masses, les capacités, les angles, les aires et les volumes dans différents contextes impliquant notamment les figures géométriques (triangles, rectangles, carrés, parallélogrammes ou disques) ou les solides (cube ou parallélépipède rectangle). Ces calculs de grandeurs peuvent impliquer la connaissance de formules ou de certaines propriétés des figures.

Géométrie et espace

- Formes à deux ou trois dimensions : connaissance des figures (carré, rectangle, losange, parallélogramme, triangle, cercle) et des solides de base (parallélépipède rectangle, cylindre, sphère et cube), de leur définition et/ou de leurs propriétés.
- Positions et droites remarquables : reconnaître ou identifier des droites parallèles ou perpendiculaires dans le plan ; reconnaître ou identifier des droites remarquables (médiannes, médiatrices, bissectrice,...) dans des figures géométriques ; reconnaître ou identifier deux objets déplacés selon une symétrie orthogonale. Connaître les repères spatiaux et la façon d'indiquer les coordonnées pour identifier la position d'un point dans un repère orthonormé.

1.2.2.1 Présentation des domaines et sous-domaines de *compétences* évaluées dans le test PASEC2019 de mathématiques de fin de scolarité primaire

Le test PASEC attribue à chaque domaine les poids suivants :

Tableau 9 : pourcentages et nombres d'items consacrés à chaque domaine et sous-domaine dans le test de mathématiques de fin de primaire de l'évaluation PASEC2019

Domaines et sous-domaine de contenu	Nombre d'items	Poids dans le test
Nombres et opérations	40	48%
Nombres décimaux et fractions	16	19%
Nombres entiers positifs	24	29%
Grandeurs et mesures	30	36%
Calcul des grandeurs	10	12%
Unités de grandeurs et mesures	20	24%
Géométrie et espace	14	17%
Formes à deux ou trois dimensions	8	10%
Positions et droites remarquables	6	7%
Total général	84	100%

Les domaines qui sont évalués dans le test PASEC en fin de cycle primaire sont :

- **Nombres et opérations** :
 - Nombres entiers positifs (ou nombres naturels) : compréhension des nombres et des propriétés des quatre opérations ; opérations sur les nombres telles additionner, soustraire, multiplier, diviser.
 - Nombres décimaux et fractions : compréhension des nombres décimaux et des fractions-nombres ; compréhension de la fraction-opérateur ; compréhension des pourcentages et de leur relation avec les fractions-opérateurs ; opérations sur les nombres décimaux et sur les fractions et opérations impliquant le calcul d'un pourcentage.
- *Grandeurs et mesures* :
 - Unités de mesures : connaissance et compréhension des unités de mesure conventionnelles telles que les longueurs, les masses, les capacités, les angles, les heures et les durées ; conversion d'unités de mesures.
 - Calcul des grandeurs : opérations impliquant des unités de mesures telles que les longueurs, les masses, les capacités, les angles, les aires et les volumes dans différents contextes impliquant notamment les figures géométriques (triangles, rectangles, carrés, parallélogrammes ou disques) ou les solides (cube ou parallélépipède rectangle). Ces calculs de grandeurs peuvent impliquer la connaissance de formules ou de certaines propriétés des figures.

➤ *Géométrie et espace :*

- Formes à deux ou trois dimensions : connaissance des figures (carré, rectangle, losange, parallélogramme, triangle, cercle) et des solides de base (parallélépipède rectangle, cylindre, sphère et cube), de leur définition et/ou de leurs propriétés.
- Positions et droites remarquables : reconnaître ou identifier des droites parallèles ou perpendiculaires dans le plan ; reconnaître ou identifier des droites remarquables (médiannes, médiatrices, bissectrice,...) dans des figures géométriques ; reconnaître ou identifier deux objets déplacés selon une symétrie orthogonale. Connaître les repères spatiaux et la façon d'indiquer les coordonnées pour identifier la position d'un point dans un repère orthonormé.

1.2.2.2 Présentation des processus cognitifs évalués dans le test PASEC2019 de mathématiques de fin de scolarité primaire

Les processus cognitifs tels que définis par le cadre de référence PASEC font référence aux démarches mentales que les élèves mobilisent pour répondre aux différentes questions du test de mathématiques.

Tableau 10 : Pourcentages et nombres d'items consacrés à chaque processus cognitif dans le test de mathématiques de fin de primaire de l'évaluation PASEC2019

Processus cognitif	Nombre d'items	Poids dans le test
Connaître et comprendre des concepts	38	45 %
Appliquer des procédures	34	41%
Résoudre des problèmes	12	14 %
Total général	84	100%

Le croisement de ces dimensions avec les domaines et sous-domaines de contenu permet d'apprécier globalement la structure du test administré aux élèves en mathématiques et la répartition souhaitée des questions sur l'échelle de résultat. En effet, cette répartition des questions vise également à estimer la répartition souhaitée du score entre ces différentes catégories puisque chaque question est corrigée de façon dichotomique.

Le tableau suivant présente la répartition des questions dans les épreuves définitives PASEC2019 selon les trois processus cognitifs et les trois domaines et six sous-domaines de contenu.

Tableau 11 : Pourcentages et nombres d'items consacrés à chaque domaine et processus cognitif dans le test de mathématiques de fin de primaire de l'évaluation PASEC2019

Domaines et sous-domaine de contenu	Connaître et comprendre des concepts	Appliquer des procédures	Résoudre des problèmes	Total général
Nombres et opérations	12	18	10	40
Nombres décimaux et fractions	6	7	3	16
Nombres entiers positifs	6	11	7	24
Grandeurs et mesures	8	16	6	30
Calcul des grandeurs	-	7	3	10
Unités de grandeurs et mesures	8	9	3	20
Géométrie et espace	14	-	-	14
Formes à deux ou trois dimensions	8	-	-	8
Positions et droites remarquables	6	-	-	6
Total général	38	34	12	84
	45 %	41 %	14%	100 %

1.2.3 Protocole et spécificités des tests PASEC2019 de fin de scolarité primaire

➤ Méthode des cahiers tournants

Cette méthode, utilisée par les principaux programmes internationaux d'évaluation des acquis des élèves, permet de diviser un test en plusieurs blocs d'items afin de faire passer un sous-ensemble du test, qui compose un livret, à une partie de l'échantillon des élèves. L'utilisation du modèle de réponse à l'item (IRT, pour *Item Response Theory*) permet de rapporter, sur une seule et même échelle et sous certaines contraintes, les performances obtenues avec différents livrets de contenus comparables. Des items d'ancrage permettent de faire un lien entre tous les livrets, de placer tous les items du test sur une même échelle et de prédire le score des élèves aux items qui ne leur ont pas été soumis.

L'ordre de passation des blocs est organisé de façon à ce que chaque bloc d'items occupe tous les rangs d'un livret à l'autre, d'où l'appellation « cahiers tournants ».

➤ Organisation des items dans les tests PASEC2019

Le test complet de compréhension de l'écrit est composé de 98 items répartis dans 4 blocs d'items distribués dans 4 livrets de test. Le test de mathématiques suit la même logique d'organisation et comprend 84 items. Les items étant répartis dans des blocs de 20 à 21 items, eux même positionnés dans un des 4 livrets (livret A/B/C/D) selon la méthode des cahiers tournants.

Chacun des blocs est réparti deux fois dans les 4 livrets (Livret A/B/C/D). Les 8 blocs (4 blocs de lecture et 4 blocs de mathématiques) sont positionnés dans les 4 livrets de manière que chaque bloc apparaisse une fois en début et une fois en fin de livret.

Tableau 12 : Répartition des blocs d'items dans les livrets

Livret A	Bloc 1 Lecture	Bloc 2 Lecture	Bloc 1 Mathématiques	Bloc 2 Mathématiques
----------	----------------	----------------	----------------------	----------------------

Livret B	Bloc 2 Lecture	Bloc 3 Lecture	Bloc 2 Mathématiques	Bloc 3 Mathématiques
Livret C	Bloc 3 Lecture	Bloc 4 Lecture	Bloc 3 Mathématiques	Bloc 4 Mathématiques
Livret D	Bloc 4 Lecture	Bloc 1 Lecture	Bloc 4 Mathématiques	Bloc 1 Mathématiques

Selon cette répartition, chaque élève enquêté répond uniquement à environ 46 items en lecture et 40 items en mathématiques.

Les 4 livrets sont distribués aléatoirement dans chaque classe à 6 ou 7 élèves maximum ; au total, 25 élèves maximum d'une même classe seront évalués dans une école.

Les items de compréhension de l'écrit sont répartis dans chacun des blocs de manière à former des blocs sollicitant des niveaux de processus comparables et comportant des supports écrits de même nature.

En mathématiques, les items sont répartis de façon équilibrée en termes de domaine et de processus entre les quatre blocs d'items.

Le centre international du PASEC attribue aléatoirement un ordre de passation des tests de compréhension de l'écrit et de mathématiques dans les écoles de telle sorte que la moitié des écoles de l'échantillon commence par le test de lecture et l'autre moitié, par les mathématiques. Ce mécanisme permet de neutraliser au niveau national les effets liés à l'apprentissage et à la fatigabilité d'un test par rapport à l'autre.

Dans chaque école, 25 élèves maximum sont sélectionnés, de façon aléatoire, pour participer à l'enquête PASEC. La collecte de données dans l'école est assurée par un administrateur de test.

L'administration des tests de compréhension de l'écrit et mathématiques s'effectue en deux jours dans chaque école et a lieu uniquement en matinée. La durée des tests de compréhension de l'écrit et de mathématiques est de 2 heures maximum chacun, avec une pause de 10 minutes après une heure.

Partie III. Test des enseignants du primaire dans le cadre de l'évaluation PASEC2019

L'évaluation PASEC2019 introduit un module supplémentaire à destination des enseignants pour mesurer et comparer au niveau international et national leur (i) niveau de maîtrise des contenus disciplinaires en compréhension de l'écrit et en mathématiques sur les composantes enseignées au primaire et sur (ii) leur niveau en didactique de la compréhension de l'écrit et des mathématiques du programme de l'enseignement primaire.

1.3.1 Les domaines disciplinaires et didactiques en compréhension de l'écrit et en mathématiques et leurs compétences évaluées du test PASEC2019 des enseignants primaires

Les domaines suivants sont ciblés dans le test :

- En compréhension de l'écrit, nous avons le seul domaine : la compréhension du texte.
- En mathématiques, les domaines suivants ont été évalués :
 - nombres et opérations,
 - grandeurs et mesures,

- géométrie et espace

Les compétences ci-dessous sont évaluées dans le test :

- L'enseignant maîtrise la compréhension de l'écrit dans la langue d'enseignement, il accède au sens de ce qu'il lit et a une connaissance des structures de la langue qui lui permet de l'enseigner comme discipline au primaire et de l'utiliser comme medium d'enseignement au primaire.
- L'enseignant maîtrise les savoirs mathématiques à enseigner sur le cycle primaire et a une capacité de raisonnement qui lui permet de résoudre des situations problèmes qui pourraient être proposées à des élèves du primaire.
- L'enseignant sait analyser une activité d'apprentissage de compréhension de l'écrit et en mathématiques, il sait en extraire les objectifs visés, faire des choix pour mettre en cohérence un objectif d'apprentissage et un exercice et il est capable d'identifier et analyser les types d'erreurs dans des productions d'élèves dans ces domaines disciplinaires.

1.3.2 Processus cognitifs de contenu et didactique de compréhension de l'écrit et de mathématiques dans l'évaluation PASEC2019 sur les enseignants

Tableau 13 : tableau de spécification final PASEC2019 de compréhension de l'écrit

	Nombre d'items	Pourcentage
Extraire une information explicite	17	33 %
Réaliser des inférences simples	16	31 %
Interpréter et combiner des informations	19	36 %
Total général	52	100 %

Tableau 14 : tableau de spécification final PASEC2019 de mathématiques

	Nombre d'items			Total général
	Nombres et opérations	Grandeurs et mesures	Géométrie et espace	
Connaître des concepts	9	6	9	24 (39 %)
Appliquer des procédures	10	9	–	19 (30, 5 %)
Résoudre des problèmes	11	8	–	19 (30, 5 %)
Total général	30 (48 %)	23 (37 %)	9 (15 %)	62 (100 %)

Tableau 15 : tableau de spécification final PASEC2019 de didactique compréhension de l'écrit /mathématiques

	Nombre d'items
Didactique de la compréhension de l'écrit	20

Didactique de mathématiques	10
-----------------------------	----

1.3.3 Protocole et spécificités des tests PASEC2019 des enseignants du primaire

➤ Administration des tests

La collecte de données dans l'école est assurée par un administrateur de test auprès de tous les enseignants présents le jour de l'enquête. Les enseignants de l'école sont sensibilisés au cours de la semaine de la collecte sur les objectifs de l'étude et sur le calendrier de travail.

L'administration de l'ensemble de l'enquête (tests et questionnaire contextuel) s'effectue sur une matinée dans chaque école. La durée des tests de lecture et de mathématiques est de 2 heures maximum chacun, avec une pause de 10 minutes après une heure.

Chaque session d'une heure de test correspond à un bloc d'items. Les blocs sont scellés par des autocollants de manière à ce que les enseignants qui répondent aux questions d'un bloc ne puissent pas répondre par anticipation aux questions de l'autre bloc.

Tableau 16 : Répartition des blocs d'items dans les livrets

Livret A	Bloc 1 Lecture	Bloc 2 Lecture	Bloc unique Didactique compréhension de l'écrit	Bloc 1 Mathématiques	Bloc 2 Mathématiques	Bloc unique Didactique Mathématiques
Livret B	Bloc 2 Lecture	Bloc 3 Lecture	Bloc unique Didactique compréhension de l'écrit	Bloc 2 Mathématiques	Bloc 3 Mathématiques	Bloc unique Didactique Mathématiques
Livret C	Bloc 3 Lecture	Bloc 4 Lecture	Bloc unique Didactique compréhension de l'écrit	Bloc 3 Mathématiques	Bloc 4 Mathématiques	Bloc unique Didactique Mathématiques
Livret D	Bloc 4 Lecture	Bloc 1 Lecture	Bloc unique Didactique compréhension de l'écrit	Bloc 4 Mathématiques	Bloc 1 Mathématiques	Bloc unique Didactique Mathématiques

Selon cette répartition, chaque enseignant enquêté répond uniquement à 2 blocs d'items en lecture, à l'ensemble des items en didactique de la compréhension de l'écrit, à 2 blocs d'items en mathématiques et à l'ensemble des items de didactique des mathématiques.

Les 4 livrets sont distribués aléatoirement dans chaque école aux enseignants dans une école.

Les items de lecture sont répartis dans chacun des blocs de manière à former des blocs sollicitant des niveaux de processus comparables et comportant des supports écrits de même nature.

En mathématiques, les items sont répartis de façon équilibrée en termes de domaine et de processus entre les quatre blocs dans les blocs (2 blocs de 15 items et 2 blocs de 16 items).

2. QUESTIONNAIRES CONTEXTUELS

Lors des enquêtes PASEC, de nombreuses informations sur les élèves, les classes, les écoles et les politiques éducatives sont collectées sur le terrain par l'intermédiaire de questionnaires soumis à différents acteurs : élèves, enseignants, directeurs et direction centrale des ministères. L'étude permet d'apprécier les caractéristiques des élèves, le niveau de répartition des ressources dans les écoles, les types de pratiques enseignantes et les représentations sociales de l'école et des pratiques scolaires des acteurs. Lorsque cela est possible, les résultats des élèves aux tests PASEC sont mis en relation avec l'environnement d'apprentissage pour donner des points de repère quant aux facteurs de réussite et orienter l'action éducative.

Les thématiques d'analyse privilégient une approche centrée sur les problématiques et défis éducatifs des pays PASEC de la CONFEMEN⁸ en s'attardant tout particulièrement aux questions d'équité au sein des systèmes.

Le cadre de référence des questionnaires PASEC vise à documenter les orientations méthodologiques des questionnaires PASEC. Ce chapitre présente l'intégralité du cadre de référence disponible sur le site www.pasec.confemen.org.

Encadré 2.1 : Guide du lecteur pour interpréter les résultats et utiliser les données des questionnaires PASEC

Les niveaux de compétence atteints par les élèves sur les échelles PASEC reflètent l'effet cumulatif de l'apprentissage formel, prévu dans le cursus scolaire classique, et informel, réalisé dans le milieu familial et la communauté locale. Dès lors, les résultats obtenus par les élèves ne reflètent pas exclusivement leurs apprentissages au sein des classes enquêtées ni l'effet d'un enseignant donné sur le rendement de ses élèves. De telles informations peuvent être recueillies en mettant en place des évaluations sur la progression des élèves dans un établissement ou dans une classe donnée en évaluant leurs capacités en début puis en fin de parcours.

Lors de l'interprétation des résultats, il convient de garder à l'esprit que le format proposé par le PASEC ne permet pas d'identifier « l'impact immédiat » d'une mesure de politique éducative sur la performance scolaire. L'angle d'interprétation des résultats se limite à l'étude descriptive du lien entre les caractéristiques contextuelles et les résultats des élèves aux tests PASEC. Toutefois, en suivant des méthodologies d'évaluation adaptées, des inférences causales peuvent être faites pour identifier l'action isolée d'une mesure éducative sur la réussite des élèves par le biais d'un contrôle strict de l'environnement. Des dispositifs tels que des évaluations longitudinales et des évaluations d'impact pourraient être menées à l'échelle d'un pays pour comprendre les conséquences d'une action éducative avant de généraliser sa pratique sur tout le territoire.

De surcroît, il serait erroné de généraliser à l'ensemble du système éducatif d'un pays les conséquences d'un phénomène observé à l'échelle d'un élève, d'une classe, d'une école ou d'une région ou vice-versa sans étudier les interactions entre plusieurs facteurs, ou encore d'inférer le rôle d'un élément de contexte sans tenir compte de la manière dont il opère dans la situation étudiée. Par exemple, il semble peu probable que la stratégie éducative d'un pays puisse être transférée de façon directe dans un autre pays présentant d'autres spécificités culturelles, socioéconomiques et politiques. Une pratique enseignante bénéfique auprès d'élèves en difficulté peut avoir l'effet inverse si elle est exercée avec des élèves en situation de réussite. De tels raccourcis peuvent mener à une utilisation abusive des résultats. Par ailleurs, les relations observées entre les facteurs et les performances des élèves ne sont pas linéaires : il existe des seuils à partir desquels les facteurs

⁸ Les pays PASEC de la CONFEMEN correspondent aux pays de la francophonie membres de la CONFEMEN et éligibles au PASEC.

peuvent avoir des relations nulles, voire inversées, ou bien le lien observé entre certaines variables et les scores peut se dissiper en présence d'autres variables.

Il convient également de garder à l'esprit que les données récoltées par l'intermédiaire des questionnaires PASEC sont en majorité déclaratives, sauf en ce qui concerne les informations relatives aux ressources matérielles et physiques de l'école, qui sont récoltées et contrôlées par l'enquêteur sur le terrain. Dès lors, il n'est pas exclu que les réponses des élèves, enseignants et directeurs ne reflètent qu'imparfaitement leurs pratiques ou leurs attentes ; par exemple, des enseignants pourraient ne pas livrer leur sentiment parce qu'ils sont influencés par la désirabilité sociale de certaines réponses.

Une étude comparative poussée du rôle des caractéristiques des systèmes éducatifs sur les performances des élèves, par exemple du lien entre la réussite scolaire et des variables nationales comme le PIB, nécessiterait la participation de nombreux pays. Dans le cadre de ce premier cycle d'évaluations, ce volet est très limité du fait du nombre réduit de pays considérés. À cet égard, les données globales sur les pays sont strictement utilisées pour illustrer le contexte de leur système éducatif et ne permettent pas d'étudier la relation entre les facteurs nationaux et la réussite scolaire.

Des analyses thématiques ciblées en fonction des intérêts propres à chaque pays pourraient être menées pour mettre en perspective les résultats des évaluations PASEC ; ces analyses qualitatives toucheraient à la fois les aspects pédagogiques, culturel et sociohistorique.

5.1. Approche globale du PASEC pour étudier le contexte éducatif

De 1991 à 2012, le PASEC a décrit le contexte éducatif au primaire et étudié les facteurs contextuels impliqués dans le processus d'apprentissage des élèves au cours d'une année scolaire. La méthodologie basée sur une évaluation des élèves de deuxième et d'avant-dernière années du primaire, en langue et en mathématiques, en début⁹ puis en fin d'année scolaire¹⁰ permettait d'égaliser le niveau de connaissance initial et les antécédents des élèves pour étudier la relation entre les caractéristiques des écoles, des classes et des élèves et leurs performances scolaires en fin d'année. La variabilité des conditions de scolarité au sein de chaque pays a permis de déceler un ensemble de facteurs contextuels fréquemment corrélés avec la performance scolaire des élèves quel que soit le système éducatif évalué. D'autre part, ces études (PASEC, 2012¹¹) ont conclu que certains facteurs avaient un effet controversé selon les matières, les niveaux scolaires et les pays évalués.

La récente réforme de la méthodologie du PASEC se traduit par l'adoption d'une approche internationale et comparative basée sur une seule période de test, soit en fin d'année scolaire¹², menée en début et en fin de scolarité primaire¹³ et l'introduction de nouveaux instruments de mesure pour apprécier le degré d'efficacité et d'équité dans les pays. Cette méthodologie ouvre de nouvelles perspectives d'étude et de réflexion sur l'environnement scolaire dans un cadre international tel que celui des pays de la CONFEMEN.

Un **système éducatif efficace** permet à tous les enfants de disposer des connaissances, compétences et attitudes attendues (fixées par les programmes scolaires) à la sortie de l'enseignement obligatoire et aux grands moments de la scolarité primaire. Dans le cadre des évaluations PASEC, un système est estimé très efficace lorsqu'il permet à tous les enfants, ou au moins à une grande majorité, d'atteindre certaines compétences de base : en début de scolarité primaire, celles qui doivent être acquises pour poursuivre le cycle avec profit, et en fin de cycle primaire, celles qui sont nécessaires pour poursuivre leur scolarité dans de bonnes conditions.

Un **système éducatif équitable** tend à réduire les inégalités de scolarisation et de réussite scolaire entre les différents profils d'élèves, entre les différents types d'écoles et entre les régions. Une juste répartition des moyens éducatifs entre les régions et entre les écoles à l'intérieur des régions est un premier pas vers cet objectif.

⁹ Dénommée « pré-test », cette évaluation se tenait le plus tôt possible dans l'année scolaire, à une période généralement comprise entre octobre et novembre.

¹⁰ Dénommée « post-test », cette évaluation se tenait le plus tard possible dans l'année scolaire, à une période généralement comprise entre mai et juin.

¹¹ PASEC (2012). *Synthèse des résultats des évaluations diagnostiques du Programme d'Analyse des Systèmes Éducatifs de la CONFEMEN*. Dakar: CONFEMEN.

¹² L'évaluation se tient dans tous les pays à une période comprise entre avril et mai de l'année scolaire ciblée pour l'enquête.

¹³ 2^e année du primaire et dernière année du primaire (6^e année).

Les questionnaires de l'évaluation internationale PASEC privilégient l'analyse des questions les plus pertinentes¹⁴ pour les pays à partir de données et d'indicateurs fiables et comparables dans l'espace et dans le temps.

L'enquête décrit le niveau de ressources et les pratiques éducatives à différents échelons du système. Les performances des élèves sont mises en relation avec les facteurs contextuels en fonction des recherches antérieures sur les déterminants de l'apprentissage et des processus d'enseignement et d'apprentissage en vigueur dans les pays.

La collecte de données contextuelles auprès d'échantillons représentatifs permet de généraliser les résultats à l'échelle d'un pays et des zones éducatives¹⁵ à l'intérieur de ce dernier.

Les items des questionnaires ont été conçus en français¹⁶ par le PASEC et ont été validés par son Comité scientifique. Les équipes nationales PASEC des 14 pays participant à la deuxième évaluation internationale PASEC2019 ont contribué à la mise en place de ces instruments de mesure. Le développement des questionnaires a suivi un processus scientifique conforme aux standards des évaluations internationales. La qualité des items a été pré-testée dans chacun des pays participants.

¹⁴ Les questionnaires PASEC privilégient une approche contextuelle adaptée aux pays puisque certains intrants éducatifs et processus de transformation leur sont spécifiques et conditionnés par le niveau de pauvreté dans les zones géographiques concernées. À titre d'exemple, le mode de fonctionnement pédagogique des écoles à « double flux » est spécifique aux pays en développement afin de soutenir la demande éducative dans un contexte de pénurie de salles de classe et d'enseignants.

¹⁵ Les zones éducatives proposées dans l'évaluation PASEC2019 regroupent plusieurs zones administratives. Le découpage en zones éducatives répond à des exigences méthodologiques liées à l'échantillonnage. La définition de ces zones a été opérée en collaboration avec les pays.

¹⁶ Langue source de conception des items.

5.2. Cadre théorique des questionnaires de contexte

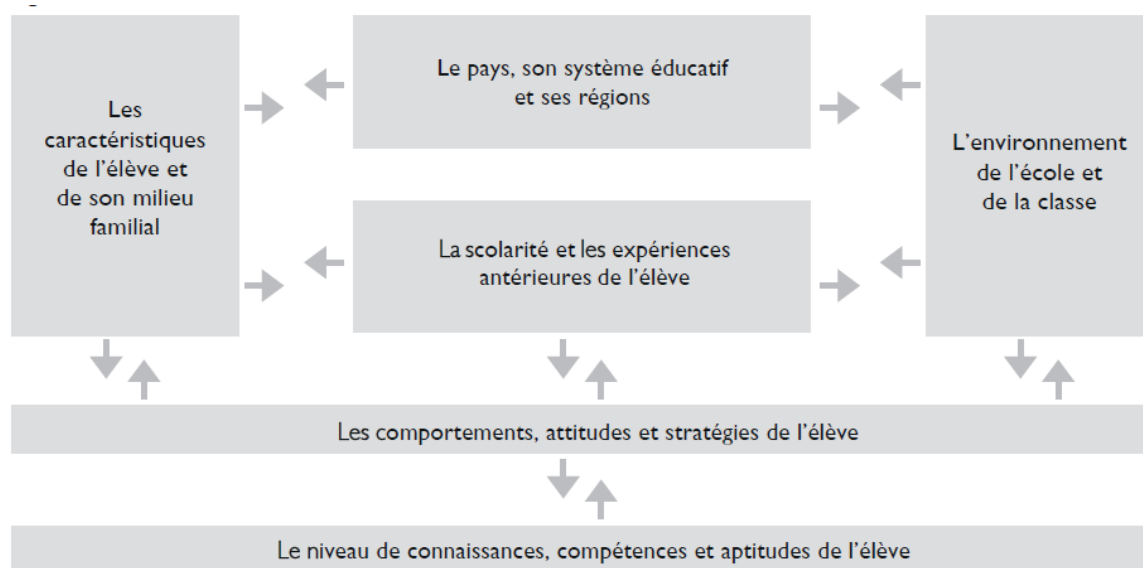
L'apprentissage des élèves à l'école est un processus complexe et multidimensionnel dans lequel interagissent l'environnement socioéconomique, le contexte éducatif, les curricula, les capacités cognitives, les attitudes, les comportements, les stratégies, les représentations et les expériences des acteurs dans le champ scolaire et dans la vie quotidienne.

Tous ces facteurs agissent directement ou indirectement sur la réussite des élèves et sur l'équité scolaire au niveau d'un pays, des zones éducatives, des écoles, des classes et enfin au niveau des élèves et de leur famille.

Par exemple, nous pouvons formuler l'hypothèse que le parcours d'un élève dans le cursus scolaire et ses expériences personnelles hors du champ de l'école sont fortement dépendants du niveau socioéconomique et culturel de sa famille. En moyenne, un élève d'une famille favorisée et alphabétisée résidant en ville et pratiquant quotidiennement la langue d'enseignement à domicile va entrer dans le système éducatif avec des compétences langagières plus développées qu'un élève d'une famille défavorisée et non alphabétisée vivant en milieu rural. De plus, un parent qui lit tous les soirs des histoires à ses enfants va participer pleinement au développement de leurs capacités de compréhension orale et de lecture ainsi que de leur goût pour la lecture.

Le schéma suivant reflète les grands champs contextuels pris en considération dans l'évaluation PASEC.

Figure 2.1 : Champs contextuels abordés dans l'évaluation PASEC2019



Dans le cadre de l'évaluation PASEC, ces différents niveaux de champs contextuels sont regroupés pour permettre une analyse statistique adaptée¹⁷ et tenir compte de la structure des données. Trois niveaux d'étude sont retenus¹⁸ :

¹⁷ Le modèle multiniveau retenu met en relation uniquement les données de niveau « élèves » et « écoles » avec les performances scolaires ; les données de niveau « pays » seront présentées seulement à titre descriptif.

¹⁸ Les données de niveau 3 ne peuvent être mises en relation avec les scores dans l'évaluation PASEC2019 en raison du nombre limité de ces ensembles.

- Niveau 1 : le milieu familial et l'élève ;
- Niveau 2 : l'école, la classe et la communauté locale ;
- Niveau 3 : le pays, son système éducatif et ses zones éducatives.

Pour chacun de ces niveaux, le PASEC cherche à étudier les moyens (ou « intrants », tels que les manuels scolaires), les processus de transformation (tels que les pratiques pédagogiques), la perception des acteurs face à leurs pratiques, les résultats et les produits du système (ou « extrants », tels que le niveau de compétence des élèves en lecture). Le cadre méthodologique ne permet pas actuellement d'étudier ces différentes dimensions de manière uniforme pour chacun des niveaux de l'échantillon (niveau « élèves », niveau « écoles » et niveau « pays »).

Dans l'ensemble, les données disponibles et les analyses menées dans le cadre du rapport international PASEC¹⁹ permettent d'étudier le contexte selon les angles d'interprétation suivants :

- La description de la répartition des ressources ou la nature des représentations et des pratiques des acteurs en fonction des différents niveaux de l'échantillon (par élève, par type d'école, par zone éducative, pour l'ensemble d'un pays ou pour tous les pays confondus). Par exemple, les données peuvent indiquer la part des élèves qui n'ont pas de manuel scolaire dans une zone éducative d'un pays donné, ou bien la part des élèves pour l'ensemble des pays dont l'enseignant déclare ne pas aider les élèves en difficulté ;
- La description de la répartition des élèves dans les différents niveaux des échelles PASEC, ou encore les différences de scores au niveau international (tous pays confondus), au niveau de chaque pays, entre les zones éducatives d'un pays, au niveau des écoles (selon les caractéristiques de l'école et de la classe) et enfin au niveau des élèves en fonction de leurs caractéristiques. Par exemple, les données peuvent indiquer les différences de scores en mathématiques entre les filles et les garçons dans un pays donné, ou bien les différences de scores entre les élèves les plus performants et les moins performants selon le statut de l'école ;
- L'analyse du lien entre les facteurs contextuels de niveau « élèves » ou « écoles » et les performances des élèves. Par exemple, les données peuvent indiquer la force du lien entre la pratique de la langue d'enseignement à domicile et les résultats des élèves en lecture, ou bien entre le fait d'être dans une classe où le niveau socioéconomique est très hétérogène et les résultats des élèves de milieu socioéconomique défavorisé.

Le caractère cyclique des évaluations PASEC permettra également d'assurer un suivi de ces tendances dans le temps. De nouvelles perspectives d'étude sont possibles d'un cycle à l'autre et permettront de cibler une thématique particulière en fonction des attentes et besoins des pays.

Dans certains cas, le PASEC a développé des indices regroupant les réponses à plusieurs questions pour faciliter l'interprétation des résultats sous la forme d'indicateurs composites.

Le tableau suivant décrit, pour chaque niveau d'études, les thématiques et les variables/indices qui leur sont associés dans le cadre de l'évaluation PASEC2019. Pour des raisons techniques, certains indices synthétiques n'ont pas pu être construits. Ces données ne sont pas exploitées dans le cadre du rapport international PASEC2019 mais restent disponibles dans les bases de

¹⁹ Les analyses secondaires menées à partir des données PASEC internationales ou nationales peuvent poursuivre un autre objectif.

données. Le lecteur est invité à consulter le chapitre 9 de ce rapport technique pour plus d'information sur le traitement des données contextuelles.

Tableau 2.1 : Variables ou indices associés aux niveaux d'études et aux thématiques de recherche

Niveau d'études	Type de questionnaire	Thématiques de recherche	Variables ou indices
Milieu familial et élève	Questionnaire Élèves	Capital socioéconomique et culturel des familles	Possession de biens matériels, caractéristiques de l'habitation, alphabétisme dans la famille, utilisation de la langue d'enseignement et de(s) langue(s) de socialisation, indice socioéconomique
		Ressources éducatives et occasions d'apprentissage à domicile	Soutien scolaire à la maison, livres et manuels scolaires à la maison, pratique de travaux extrascolaires
		Caractéristiques personnelles des élèves	Genre, âge, composition de la famille
		Parcours scolaire des élèves	Nombre de redoublements et classes redoublées, fréquentation du préscolaire, absentéisme
		Conditions physiques pour apprendre	Audition, vue, fatigue, faim, examen médical, déparasitage, utilisation de moustiquaire
		Perception de l'école et goût d'apprendre	Perception de l'école, perception de soi dans l'apprentissage de la lecture et des mathématiques, goût et intérêt pour la lecture et pour les mathématiques
École, classe et communauté locale	Questionnaires Maîtres et Directeurs	Infrastructures, équipements, fonctionnement et ressources de la classe	Niveau socioéconomique moyen des élèves de la classe, ressources matérielles et éducatives, mode de fonctionnement de la classe, taille de la classe
		Absentéisme et occasions d'apprentissage en classe	Gestion de l'absentéisme des enseignants, perception de l'absentéisme par les enseignants, retard et absentéisme des enseignants et raisons, absentéisme des élèves et raisons, durée de l'année scolaire, abandon des élèves, grèves
		Caractéristiques personnelles et professionnelles des maîtres	Genre, âge, durée et nature de la formation initiale et continue, statut, rémunération et avantages sociaux, aspirations, ancienneté, absentéisme
		Pratiques d'enseignement et représentations des maîtres	Contenus et méthodes d'enseignement en langue et en mathématiques, discipline et sanction des élèves, raisons d'abandon des élèves, conditions de travail, harcèlement
		Infrastructures, équipements, fonctionnement,	Zone d'implantation, statut de l'école, fonctionnement de l'école, ressources matérielles (cantines, sanitaires...), ressources

Niveau d'études	Type de questionnaire	Thématiques de recherche	Variables ou indices
		ressources et contrôle de l'école	éducatives (fournitures, programmes...), suivi et contrôle de l'inspection
		Ressources et implication de la communauté locale	Infrastructures et équipements de la localité, relations avec les parents d'élèves, relations et projets avec la communauté, financement de l'école, projets-école
		Caractéristiques personnelles et profil professionnel du directeur	Genre, âge, durée et nature de la formation initiale et continue, statut, rémunération et avantages sociaux, aspirations, ancienneté, absentéisme
		Pratiques administratives et pédagogiques et représentations du directeur	Management et contrôle administratif ou pédagogique au niveau des élèves et des enseignants
		Pratiques de soutien scolaire dans l'école et représentations des acteurs	Perception des difficultés des élèves par les directeurs et les enseignants, dispositifs d'encouragement et d'accompagnement, soutien scolaire à l'école

Les informations sur le milieu familial sont collectées par l'intermédiaire des questionnaires Élèves. La collecte d'informations auprès d'élèves en début de scolarité se révèle très limitée dans le cadre d'une enquête internationale, si bien que les informations contextuelles prélevées auprès des élèves de 2^e année sont réduites par rapport à celles qui sont collectées auprès de leurs camarades scolarisés en 6^e année.

L'agencement actuel de l'enquête PASEC dans les écoles ne permet pas de disposer d'informations directes sur les parents ni sur les communautés locales. Ces données sont collectées à travers les questionnaires Élèves, Maîtres et Directeurs.

Les informations sur les écoles et les classes sont basées sur les déclarations des directeurs et des enseignants en ce qui concerne leurs caractéristiques et leurs pratiques. Les données en lien avec les ressources matérielles et éducatives sont collectées par l'enquêteur avec l'appui du personnel de l'école afin de s'assurer de la présence effective de ces moyens dans l'école.

Les données qui concernent les élèves et les écoles sont disponibles dans la base de données internationale de l'évaluation PASEC2019.

Les questionnaires Élèves, Maîtres et Directeurs sont disponibles sur le site du PASEC : www.pasec.confemen.org.

Les informations sur le pays et son système éducatif sont collectées auprès des ministères de l'Éducation et par la collecte d'informations externes et comparables auprès d'organismes internationaux.

5.3. Spécifications sur les questionnaires de contexte

Les questionnaires sont disponibles sur le site du PASEC : www.pasec.confemen.org.

Questionnaire Élèves de début de scolarité (2^e année)

Le questionnaire destiné aux élèves de 2^e année est très bref du fait du jeune âge des élèves et de leurs difficultés à comprendre et à répondre à des questions complexes sur leur environnement. Il comprend des questions sur les caractéristiques familiales et sur le parcours scolaire des élèves. La passation du questionnaire dure approximativement 15 minutes par élève et se déroule de façon individuelle, un enquêteur pour un élève. L'enquêteur pose les questions à l'élève oralement, dans la langue d'enseignement ou sa langue maternelle, pour une meilleure compréhension des questions. Les réponses sont directement reportées par l'enquêteur.

Questionnaire Élèves de fin de scolarité (6^e année)

Le questionnaire destiné aux élèves de 6^e année comprend des items sur les caractéristiques familiales, sur le parcours scolaire des élèves, sur les ressentis des élèves à propos de leur bien-être à l'école, sur l'appréciation de leur travail et leur goût pour la lecture et les mathématiques, sur leur santé et sur les ressources éducatives et les occasions d'apprentissage à domicile. La passation du questionnaire dure approximativement 45 minutes par classe et se déroule de façon collective, c'est-à-dire que l'enquêteur lit les questions et les modalités de réponse dans langue d'enseignement ou la langue maternelle des élèves pour une meilleure compréhension. Les réponses sont directement reportées par les élèves avec l'appui de l'enquêteur, qui les accompagne dans cet exercice et vérifie que toute la classe progresse en même temps.

Questionnaire enseignants et questionnaire Directeurs

Le questionnaire destiné aux enseignants est composé de deux parties qui concernent, d'une part, des données déclaratives sur l'enseignant et sa classe et, d'autre part, des données explicites sur le niveau d'infrastructure ainsi que les ressources éducatives de la classe. La première partie est remplie de façon autonome par l'enseignant alors que la deuxième partie se fait en collaboration avec l'enquêteur.

Le questionnaire destiné au directeur suit la même organisation que ceux destinés aux enseignants, si ce n'est que les thèmes abordés sont spécifiques à l'école et à la fonction de directeur.

3. PLAN D'ÉCHANTILLONNAGE

Le présent chapitre traite sur la méthode d'échantillonnage du PASEC lors de l'évaluation internationale PASEC2019. Il annonce la stratégie de sélection des écoles, des classes et les élèves dans les pays participant à l'évaluation de façon à respecter les standards d'échantillonnage retenus par le PASEC et leur mise en œuvre tout au long du processus de l'enquête, garantissant ainsi la validité et la comparabilité des résultats au niveau international. Le chapitre comporte trois grandes parties : (i) la méthodologie d'échantillonnage globale du PASEC ; (ii) l'échantillonnage des écoles dans les pays ; (iii) l'échantillonnage des classes et des élèves au sein des pays.

3.1 Procédures d'échantillonnage du PASEC2019

Comme au PASEC2014, l'enquête internationale PASEC2019 cible principalement l'ensemble des élèves de début (2^e année) et de fin (5^e/6^e année) du primaire, quel que soit le type d'école (école publique, privée...) et le milieu (Rural/urbain) d'implantation des écoles. Les données sont collectées à partir d'un échantillon représentatif au niveau national des écoles primaires d'un pays présentant les deux niveaux d'enseignement considérés. Le choix stratégique d'échantillonnage est fondé sur les contraintes en ressources humaines qualifiées et compétentes, en ressources financières capable de couvrir l'ensemble des activités, en capacité technique de gestion des données en un temps limité, etc.

L'échantillon PASEC permet d'estimer avec une certaine précision les résultats scolaires des pays participant à l'évaluation sans avoir à enquêter tous les élèves scolarisés. Des normes et des mécanismes de contrôle de la qualité sont mises en place par le programme tout au long du processus de l'évaluation pour garantir la fiabilité des échantillons et la comparabilité des résultats entre les pays et dans le temps.

La procédure d'échantillonnage en lien avec la sélection des écoles est conduite par le PASEC en collaboration avec les pays. Les écoles sont échantillonnées dans chaque pays à partir d'une base de sondage des écoles, la plus récente et qui fournit des informations détaillées sur les écoles. Les écoles ont été sélectionnées dans chaque pays selon une procédure standardisée : un échantillonnage systématique proportionnel à la taille cumulée des effectifs des élèves de début (2^e année) et de fin (5^e /6^e année) primaire. Ainsi, la probabilité de sélectionner une école de grande taille est plus grande que celle des petites écoles. À l'inverse, la probabilité d'être sélectionné à l'intérieur de l'école (intra-école) est plus faible pour un élève d'une école à grand effectif que pour celui d'une école à petite taille.

La taille standard de l'échantillon d'écoles de l'enquête PASEC2019 pour l'évaluation des élèves de fin d'année primaire est de 180 écoles. Cette taille d'échantillon est choisie pour obtenir une précision équivalente à celle qu'on obtiendrait avec un échantillon aléatoire et simple de 400 élèves²⁰. Dans certains cas, un échantillon d'écoles plus important est sélectionné lorsqu'un pays souhaite enquêter sur des problématiques éducatives spécifiques. Dans ce cas, le PASEC peut suréchantillonner les écoles de certaines des zones de ce pays afin de disposer d'un échantillon conséquent permettant de désagréger les résultats jusqu'au niveau du sous-groupe concerné. Cela a été le cas pour le Burkina Faso (400 écoles échantillonnées), le Cameroun (280 écoles) et le Togo (280 écoles). Ces différences de

²⁰ Résultat obtenu sous l'hypothèse que le coefficient de corrélation intra-classe a une valeur maximale de 0,40.

taille d'échantillon n'invalident pas la méthodologie de l'évaluation et permettent au contraire d'améliorer la précision des estimations par rapport à un échantillon standard de 180 écoles.

L'échantillon des écoles pour l'évaluation des élèves de fin primaire (5^e/6^e année) est le premier qui est tiré. Sur la base de cet échantillon, un sous-échantillon (dont la taille correspond au moins à la moitié de celui tiré en fin de primaire) est sélectionné pour l'évaluation de début (2^e année) de primaire selon un tirage aléatoire simple, si bien que chaque école sélectionnée pour les tests de fin de scolarité a une chance sur deux d'être également enquêtée en début de scolarité. Pour les pays dont l'échantillon principal est de 180 écoles, l'échantillon de début de primaire est donc constitué de 90 écoles.

L'enquête PASEC2019 a procédé à un échantillonnage en trois étapes :

- La première étape a consisté en la sélection de 180 écoles selon une procédure systématique et une probabilité proportionnelle au nombre d'élèves inscrits en début (2^e) et fin (5^e/6^e) années d'enseignement au primaire. Ces 180 écoles sont aléatoirement extraites d'une liste (base de sondage officielle pour l'année scolaire 2017-2018) qui tient compte de l'effectif des élèves de classes des niveaux d'enseignement concernés de l'ensemble des écoles du pays. Le PASEC a procédé à l'échantillonnage parmi les écoles ayant au moins une classe en fin du primaire.
- La difficulté liée à la passation individuelle des tests en en début de primaire et la nécessité d'harmoniser les pratiques entre les administrateurs de tests ont conduit le PASEC à réduire la taille de l'échantillon des écoles. Au sein de chacune des strates, seule la moitié des écoles sélectionnées aléatoirement pour l'évaluation en fin de scolarité a été invitée à participer à l'évaluation en début de scolarité. L'échantillon de 2^e année se limite donc à 90 écoles.
- Au sein de chacune de ces écoles sélectionnées, une classe de fin du primaire (5^e/6^e année) est sélectionnée parmi l'ensemble des classes du même niveau d'enseignement selon une procédure aléatoire simple. Cette procédure est réitérée au niveau de la 2^e année si l'école figure dans le sous-échantillon d'écoles qui participent à l'évaluation en début de scolarité.

3.1.1 Couverture de la population cible et taux de participation au niveau des écoles et des élèves

3.1.1.1 *Couverture de la population cible internationale*

La population cible à enquêter comprend tous les élèves inscrits dans le niveau enquêté dans une école dont la langue d'enseignement officielle est la langue utilisée pour le test. Certaines contraintes nationales peuvent conduire à réduire la couverture de la population cible en excluant tous les élèves d'une école ou certains élèves dans une école pour les raisons suivantes : difficultés d'accès à l'école,

conflit armé, très petite école²¹, élèves avec des dysfonctionnements intellectuels ou physiques permanents, etc.

Le PASEC veille à ce que les exclus de la base de données ne dépasse guère 5 % des élèves inscrits dans le niveau concerné. Les différentes exclusions au sein de la population cible peuvent s'opérer au moment de la définition de la population cible (réduction du recouvrement de la population cible) et lors de la collecte des données.

Au moment de la collecte des données dans les écoles, les élèves qui présentent des dysfonctionnements intellectuels ou physiques permanents les empêchant de répondre aux tests ou au questionnaire sont également exclus de l'enquête. Cependant, dans tous les pays du PASEC2019, les règles suivantes ont été respectées :

- un élève faible mais qui ne présente aucun dysfonctionnement intellectuel ne peut être exclu sur cette base ;
- un élève avec un handicap temporaire l'empêchant de répondre à une partie du test ou à un questionnaire ne doit pas être exclu. Par exemple, un élève avec un bras cassé ne sera pas exclu, mais il sera considéré comme absent s'il n'est pas disponible le jour de l'enquête. Il ne sera donc pas remplacé ;
- un élève avec un dysfonctionnement moteur permanent est exclu.

3.1.1.2 Taux de participation au niveau des écoles et des élèves

La participation des écoles et des élèves permet de contrôler la validité de la représentativité de l'échantillon par rapport à la population cible.

3.1.2 Taux de participation au niveau des écoles

À la suite de la mise à disposition de l'échantillon des écoles sélectionnées pour l'évaluation des systèmes éducatifs des pays, les contraintes opérationnelles peuvent conduire à ne pas enquêter une école et les élèves qui y sont scolarisés. Dans ce cas, le pays peut remplacer l'école initiale par une autre école de caractéristiques similaires préalablement identifiée pour servir d'école de remplacement. Le PASEC met à la disposition de chaque pays une liste d'écoles de remplacement choisies selon les mêmes procédures que les écoles de l'échantillon de départ. Chaque école sélectionnée dispose d'au moins une école de remplacement et au maximum de deux écoles de remplacement.

Le motif pour ne pas enquêter une école doit être spécifiée par l'administrateur.

Lors de la collecte des données du PASEC2019, les pays ont été autorisés à procéder à des remplacements d'écoles lorsque celles-ci sont momentanément fermées (épidémie, élection, examen,

²¹ Les très petites écoles peuvent être exclues pourvu que leurs effectifs combinés ne représentent pas plus de 1 % de la population d'élèves.

formation des enseignants, grève, événements particulier), situées dans un lieu inaccessible (inondations, problèmes de sécurité, route impraticable) ou par refus du chef d'établissement malgré les tentatives faites pour le ramener à la raison.

En revanche, une école n'est pas remplacée lorsqu'elle ne compte pas d'élèves dans le niveau d'enseignement à enquêter ou qu'elle est définitivement fermée. Dans ce cas, l'école est définitivement perdue. Le PASEC travaille en étroite collaboration avec les pays afin de disposer de bases de sondage fiables et à jour pour éviter les cas d'écoles définitivement fermées ou ne disposant pas de classe de début ou de fin primaire concernée par l'enquête.

Le PASEC fixe le seuil minimum de participation des écoles à 85 % du total des écoles initialement sélectionnées pour assurer la représentativité de l'échantillon par rapport à la population cible. Ce taux est calculé avant et après la prise en compte des écoles de remplacement. Le taux de participation des écoles avant la prise en compte des écoles de remplacement se calcule selon la formule suivante :

$$TxPE_{avr} = (\text{nombre d'écoles participantes dans le pays} / \text{nombre d'écoles prévues par l'enquête}) \times 100$$

Lorsque le taux de participation des écoles est inférieur à 85 % avant la prise en compte des écoles de remplacement, le taux de participation exigé après remplacement des écoles doit atteindre au minimum 85 % pour autant que le taux initial soit supérieur à 65 %.

Le taux de participation des écoles après la prise en compte des écoles de remplacement se calcule selon la formule suivante :

$$TxPE_{apr} = 85 + (85 - \text{taux initial avant remplacement}) / 2$$

Par exemple, si le taux de participation des écoles initialement échantillonnées s'élève à 82,8 % avant la prise en compte des écoles de remplacement, alors le pays devra atteindre un taux de participation après remplacement des écoles de $85 + (85 - 82,8) / 2 = 86,1$ %.

Pour que les données d'un pays soient validées et acceptées pour publication, le taux de participation des écoles doit donc être au moins égal à 85 % du nombre total d'écoles prévu dans l'échantillon prévisionnel. Une école est considérée comme non participante lorsque moins de 25 % des élèves échantillonnés pour le niveau considéré ont participé aux tests.

Les taux de participation des écoles sont calculés sans et avec pondération. Les procédures pour le calcul des taux de participation pondérés et non pondérés sont présentées au chapitre 6 de ce rapport.

3.1.3 Taux de participation au niveau des élèves

Au niveau des élèves, le taux de participation doit atteindre au moins 80 % du nombre total d'élèves échantillonnés pour un pays donné. Un élève est considéré comme participant au sein des écoles s'il a participé à l'une des quatre sessions de test : le test de lecture et le test de mathématiques sont composés chacun de deux sessions.

Les données d'un pays sont validées sans aucune réserve lorsque le pays satisfait aux normes recommandées de taux de participation.

3.2 Procédures techniques d'échantillonnage des écoles pour les pays

La validité des résultats de l'évaluation internationale PASEC2019 implique d'harmoniser l'ensemble des procédures de l'enquête notamment l'échantillonnage dans tous les pays participants. En ce qui concerne la phase d'échantillonnage, tous les pays ont suivi de façon standardisée la même procédure pour la définition de la population cible nationale, la construction de la base de sondage, la définition des variables de stratification, l'échantillonnage des écoles et la sélection des classes et des élèves dans les écoles. Chacune de ces étapes est détaillée dans les sections suivantes.

3.2.1 Définition de la population cible nationale

La définition de la population cible a été réalisée pour chaque pays par le PASEC en collaboration avec les équipes nationales PASEC des pays participants. Ces opérations sont réalisées à partir de la base de sondage nationale et peuvent mener à l'exclusion des écoles de zones spécifiques (conflit armé, inaccessibilité) avant le tirage des écoles. Lorsque des difficultés risquent de survenir dans des zones spécifiques, le PASEC a échangé avec les pays afin de tenir compte de cette réduction de couverture dans le processus et de ne pas dépasser le seuil maximal d'exclusion fixé par le PASEC.

Dans les cas de réduction de la couverture de l'enquête, les pays du PASEC2019 ont été invités à spécifier les raisons d'exclusion des écoles de la base de sondage envoyée au PASEC.

Les pays ont été sensibilisés aux cas d'exclusion d'élèves et de remplacement ou non des écoles. Deux fiches de suivi écoles et élèves ont été élaborées et mises à la disposition des pays pour que les administrateurs puissent en disposer lors de l'enquête.

3.2.2 Élaboration de la base de sondage

Lors de la préparation de la base de sondage, il a été demandé aux pays de :

- **dater les données** : il s'agit de mentionner l'année académique couverte par la base de sondage. Chaque pays doit fournir les données les plus récentes dont il dispose et tout en privilégiant la qualité de ces données. Par exemple, il est préférable de disposer de données vérifiées de l'année 2018 que de données de 2019 en cours de nettoyage. Toutefois, le souhait est d'avoir des données fiables, vérifiées et validées de 2019 ;
- **documenter les exclusions** : une réduction de la couverture de la population cible doit être documentée. Les taux de population concernés sont calculés et comparés aux normes recommandées par le PASEC ;
- **définir les variables de stratification** : le PASEC a défini avec les pays les variables pertinentes pour la stratification. Lors de la constitution de la base de sondage, les pays s'assurent que les variables de stratification sont disponibles pour chaque école contenue dans la base. Il est par exemple inutile d'envisager de stratifier un échantillon sur la base du type d'école si l'information manque pour chaque école dans la base de données transmise au PASEC ;

- **attribuer un identifiant unique à chaque école** : les pays utilisent généralement une identification nationale pour reconnaître une école. Dans le cas où un identifiant national n'existe pas, un numéro d'identification unique est attribué à chaque école, deux écoles différentes ne pouvant avoir le même identifiant. En plus de l'identifiant, le nom de l'école, l'adresse, un numéro de téléphone s'il existe et le nom d'une personne de contact sont mentionnés.

Une base de sondage qui répond aux normes PASEC respecte les critères suivants :

- unicité des identifiants des écoles ;
- unicité de l'adresse des écoles ;
- inexistence de données manquantes pour les variables de taille des classes de 2^e et de 5^e/6^e années, les variables de stratification, l'identifiant et l'adresse de l'école ;
- concordance entre le nombre d'écoles indiqué dans la base de sondage et le nombre d'écoles attendu ;
- cohérence entre la taille des deux populations d'élèves (somme des élèves de 5^e/6^e année du primaire et somme des élèves de 2^e année) avec les statistiques générales du pays.

Précisions sur la mesure de la taille des écoles : L'enquête PASEC2019 cible deux populations distinctes. Deux variables de taille sont donc attendues : le nombre d'élèves inscrits en fin primaire (5^e / 6^e année) et le nombre d'élèves inscrits en début primaire (2^e année). Dans le cas où ces informations ne sont pas disponibles avec ce niveau de précision, le pays est invité à fournir, pour chaque école, le nombre total d'élèves et le nombre de niveaux scolaires organisés. La taille est alors estimée en divisant le nombre total d'élèves par le nombre de niveaux scolaires organisés. Une école pour laquelle aucune information de taille n'est communiquée n'est pas prise en compte pour tirer un échantillon proportionnel à la taille. Il est donc indispensable de mettre tout en œuvre pour fournir une mesure de taille pour chacune des écoles du pays. Le même type de mesure de taille est utilisé pour toutes les écoles d'un pays donné. La mesure de taille ne peut être le nombre d'élèves par niveau pour certaines écoles et pour d'autres une approximation au départ de la population totale. Tous les pays participant à l'évaluation PASEC2019 ont fourni la même mesure de taille selon le nombre d'élèves inscrits en classe de début et de fin primaire.

3.2.3 Définition des variables de stratification

Dans le but d'augmenter la précision de l'échantillon, toutes les écoles de la base de sondage sont organisées en groupes homogènes selon une ou plusieurs variables déterminantes, dites « variables de stratification ». La stratification est appliquée au premier niveau d'échantillonnage, c'est-à-dire au niveau des écoles. Deux types de stratification sont utilisés dans cette phase : la stratification explicite et la stratification implicite.

La stratification explicite consiste en la création de groupes d'écoles ou de strates homogènes et distinctes selon la variable de stratification explicite définie. L'échantillonnage des écoles est effectué de manière indépendante dans chacune des strates ainsi formées, et le nombre d'écoles sélectionnées par strate est généralement proportionnel à la taille relative à l'effectif total des élèves inscrits dans la strate. Plus la variable de stratification corrèle avec la variable d'intérêt de l'étude, plus les échantillons de chacune des strates seront homogènes vis-à-vis de la performance et plus la variance d'échantillonnage sera réduite.

Les variables de stratification explicite utilisées dans le cadre du PASEC2019 portent en général sur les regroupements de régions, de départements ou de provinces en fonction de la structure administrative du pays. 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. Les variables de stratification implicite utilisées sont le type d'école (public, privé ou communautaire) et le milieu d'implantation (urbain ou rural). Cette stratification implicite est utile pour améliorer la représentativité de l'échantillon suivant ces variables définies.

Le tableau ci-dessous présente pour chaque pays les variables de stratification (explicite et implicite) utilisées ainsi que le nombre de strates définies. À titre d'exemple, le Burkina Faso compte 12 régions qui ont été regroupées pour former 8 strates. Le type d'école (public ou privé) a été utilisé comme variable de stratification implicite.

Tableau 3.1 : Variables de stratification utilisées pour chaque pays

Pays	Variables de stratification explicite	Nombre de strates explicites	Variables de stratification implicite
Bénin	Département (12)	12	Type d'école (2)
Burkina Faso	Région (13)	13	Type d'école (2)
Burundi	Province (18)	18	Type d'école (2)
Cameroun Francophone	Région (10)	8	Type d'école (2)
Cameroun Anglophone	Région (10)	4	Type d'école (2)
Congo	Département (12)	12	Type d'école (2) et Milieu d'implantation (2)
Côte d'Ivoire	Région (31)	6	Type d'école (2)
Gabon	Région (9)	9	Type d'école (2) et Milieu d'implantation (2)
Guinée	Région (8)	8	Type d'école (2) et Milieu d'implantation (2)
Madagascar	Région (22)	22	Type d'école (2) et Milieu d'implantation (2)
Niger	Région (8)	8	Type d'école (2) et Milieu d'implantation (2)
RDC	Région (11)	11	Type d'école (2) et Milieu d'implantation (2)
Sénégal	Région (14)	5	Type d'école (2) et Milieu d'implantation (2)
Tchad	Région (22)	22	Type d'école (3) et Milieu d'implantation (2)
Togo	Région (6)	6	Type d'école (2) et Milieu d'implantation (2)

3.2.4 Attribution d'une mesure de taille à chaque école

Pour rappel, la méthode de tirage des écoles (systématique proportionnelle à la taille combinée des effectifs de début et de fin du primaire) implique qu'une mesure de taille (MOS, pour *Measure of size*) soit calculée pour chaque école de la base de sondage. La taille combinée d'une école est obtenue en additionnant le nombre d'élèves de début d'année et le nombre d'élèves de fin de scolarité primaire de l'école. La taille par niveau scolaire (SIZEGRAD) est ensuite calculée en divisant la taille combinée par deux.

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. Le plan d'échantillonnage (pour l'échantillon standard) prévoit qu'un seul échantillon de 180 écoles soit tiré avec une probabilité proportionnelle à la taille de chaque école. Ensuite, parmi ces 180 écoles sélectionnées pour la fin d'année du primaire, 90 écoles sont à nouveau sélectionnées pour l'évaluation des élèves de la 2^e année du primaire.

La MOS est égale à la taille estimée par niveau scolaire (SIZEGRAD) pour autant que celle-ci soit supérieure ou égale à la taille théorique de l'échantillon par classe²². Dans le cas contraire, la MOS est égale à 25. Formellement, $MOS = \max(SIZEGRAD, 25)$. Lorsque la taille estimée par niveau scolaire (SIZEGRAD) est plus grande que 25, alors la MOS sera égale à la taille estimée par niveau scolaire (SIZEGRAD).

3.2.5 Calcul des taux d'exclusion d'écoles

À partir de la base de sondage nettoyée de toute incohérence, les informations sur les exclusions éventuelles sont données en termes de pourcentage d'élèves concernés. Pour garantir la représentativité de l'échantillon, le taux d'exclusion ne doit pas dépasser le seuil de 5 % par niveau scolaire quel que soit le type d'exclusion. Lorsque le taux d'exclusion respecte le seuil établi, les écoles sont définitivement exclues de la population effective. La base de sondage est alors nettoyée de ces exclusions.

3.2.6 Traitement des petites écoles et des très petites écoles

Le PASEC définit une très petite école comme une école qui compte moins de 8 élèves pour un niveau cible. Les très petites écoles sont exclues lorsque, au total, le nombre total d'élèves inscrits pour l'ensemble des très petites écoles ne dépasse pas 1 % de la population cible. Lorsque les très petites écoles ne peuvent être exclues parce que le pourcentage d'élèves qu'elles comptent est supérieur à 1 %, une strate est créée pour ces très petites écoles.

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 25 élèves en dernière année du primaire. Une grande école est une école fréquentée par au moins 26 élèves en fin du primaire.

Si le pourcentage correspondant à l'ensemble des élèves inscrits dans des petites écoles est égal ou supérieur à 4 % de la population cible, 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 proportionnelle à la taille.

²² En 6^e année, 25 élèves doivent passer les tests prévus pour l'évaluation.

3.2.7 Échantillonnage des écoles

La définition des variables de stratification permet de répartir les écoles dans les différentes strates prévues. Le nombre total d'écoles à échantillonner dans chaque pays est réparti entre les différentes strates selon le poids de la strate. Les étapes suivantes sont réalisées :

3.2.7.1 Tri des écoles dans la base de sondage

Les écoles sont triées selon la variable MOS au sein de chaque strate explicite et implicite en alternant suivant un tri ascendant et un tri descendant des observations. De façon pratique, les tris sont effectués d'abord selon la première variable de stratification, puis selon la deuxième variable de stratification et enfin selon la MOS.

3.2.7.2 Processus de sélection des écoles

Suite au tri des écoles, un pas d'échantillonnage (INT) est défini pour chaque strate explicite en divisant le nombre total d'élèves du niveau enquêté dans la strate (N_h) par le nombre d'écoles à échantillonner dans la strate (n_{e_s}). Formellement, on a :

$$INT = \frac{N_h}{n_{e_s}}$$

Où :

$$N_h = \sum_{i=1}^{N_{e_s}} MOS_i$$

Le calcul du pas d'échantillonnage permet de détecter dans la base de sondage les très grandes écoles appelées « 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 à 1. La détection des écoles certaines se fait à l'intérieur de chaque strate explicite initiale puisque les pas d'échantillonnage varient d'une strate à l'autre.

Un nombre aléatoire (NA) issu d'une distribution uniforme $[0;1]$ est tiré avant de procéder à la sélection des écoles. Cette étape confère le caractère aléatoire de la sélection de l'échantillon d'écoles. Ce nombre aléatoire est multiplié par le pas d'échantillonnage arrondi à l'unité supérieure. Le produit obtenu (EST_1) permet d'identifier la première école à sélectionner ($EST_1 = NA * INT$). L'école qui contient l'élève $E_i = EST_1$ est alors sélectionnée, les i élèves de la liste d'écoles triées étant préalablement numérotés de 1 à N_h .

La deuxième école à sélectionner est obtenue en ajoutant le pas d'échantillonnage au nombre EST_1 de telle sorte que $EST_2 = EST_1 + INT$. Cette procédure systématique est répétée jusqu'à ce que le nombre d'écoles à échantillonner soit atteint. Par exemple, la troisième école sélectionnée est obtenue en calculant d'abord $EST_3 = EST_1 + INT + INT$. L'école qui contient l'élève $E_i = EST_3$ est alors sélectionnée.

Pour illustrer les différentes étapes de la méthode de tirage proportionnelle à la taille, nous supposons qu'à partir de la base de sondage d'un pays disposant d'une seule strate, le nombre total d'élèves est

de 420 et le nombre total d'écoles est de 10. Nous souhaitons sélectionner quatre écoles dans cette strate.

Encadré 3.1 : Illustration du tirage proportionnel à la taille de 4 écoles dans une strate de 10 écoles et de 420 élèves

- Calcul du pas d'échantillonnage : $INT=420/4=105$
- Tirage d'un nombre aléatoire : $NA=0,232$
- Le premier nombre sélectionné sera $0,232 * 105 = 24,36$ dont l'arrondi supérieur est 25. Ce premier nombre sélectionné permet de déterminer la première école échantillonnée. Ainsi, l'école A002 est sélectionnée puisqu'elle contient l'élève 25.
- Le deuxième nombre sélectionné est $25 + 105 = 130$. Ce second nombre permet de déterminer la deuxième école échantillonnée, soit l'école A006 puisqu'elle contient l'élève 130.
- Le troisième nombre sélectionné est $105 + 130 = 235$. Ce troisième nombre permet de déterminer la troisième école échantillonnée soit l'école A008 puisqu'elle contient l'élève 235.
- Le dernier nombre sélectionné est $105 + 235 = 340$. Ce dernier nombre permet de déterminer la quatrième école échantillonnée, soit l'école A010 puisqu'elle contient l'élève 340.

Ces quatre écoles constituent alors l'échantillon d'écoles.

Identifiant de l'école	Mesure de taille de l'école au niveau cible (MOS)	MOS cumulée	Numéro E_i des élèves		École échantillonnée
			de	à	
A001	20	20	1	20	
A002	20	40	21	40	Sélectionnée
A003	20	60	41	60	
A004	25	85	61	85	
A005	30	115	86	115	
A006	35	150	116	150	Sélectionnée
A007	40	190	151	190	
A008	45	235	191	235	Sélectionnée
A009	85	320	236	320	
A010	100	420	315	420	Sélectionnée

3.2.7.3 Désignation des écoles de remplacement

Pour éviter une réduction de taille de l'échantillon due aux circonstances de la collecte des données sur le terrain, un maximum de deux écoles de remplacement est prévu pour chaque école échantillonnée initialement. Dans la base de sondage ordonnée selon la taille de l'école et au sein de chaque strate, l'école qui suit directement l'école échantillonnée est désignée comme sa première école de remplacement (R1) et l'école qui la précède est sa seconde école de remplacement (R2). Il

peut y avoir des cas où l'école initialement échantillonnée ne dispose que d'une seule école de remplacement. 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 d'un autre refus ou de toute autre circonstance empêchant la collecte des données auprès de la première école de remplacement.

3.2.7.4 Sélection des écoles de l'échantillon de début de scolarité (2^e année) du primaire

La difficulté liée à la passation individuelle des tests en début de scolarité primaire et la nécessité d'harmoniser les pratiques entre les administrateurs de tests ont conduit le PASEC à réduire la taille de l'échantillon des écoles de ce niveau. Au sein de chacune des strates constituées préalablement, seule la moitié des écoles sélectionnées aléatoirement pour l'évaluation en fin de scolarité a été invitée à participer à l'évaluation en début de scolarité.

3.2.7.5 Attribution d'un identifiant PASEC à chaque école

Un identifiant PASEC (ID_PASEC) est attribué à chaque école de l'échantillon ainsi qu'à chacune des écoles de remplacement, en dehors de l'identifiant national fourni lors de la transmission de la base de sondage par le pays. L'identifiant PASEC est construit de manière à reconnaître facilement les écoles échantillonnées dans chaque strate explicite et pour chaque niveau scolaire. Cet identifiant unique est formé de six chiffres :

- deux chiffres permettant d'identifier la strate : les strates explicites sont numérotées séquentiellement (01, 02, 03, etc.) ;
- un chiffre permettant d'identifier le ou les niveaux évalués : « 1 » si l'école est sélectionnée pour la fin du primaire (5^e/6^e année) uniquement, « 2 » si elle fait partie des deux échantillons (début et fin d'années primaire) ;
- trois chiffres permettant d'identifier l'école :
 - Au sein de chaque strate, les écoles sélectionnées (S) pour l'échantillon principal (5^e / 6^e année du primaire) sont numérotées séquentiellement à partir de 001. Par exemple, si 80 écoles sont sélectionnées dans une strate, les écoles sélectionnées 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 initialement sélectionnée (S) correspondante. Il s'agit de l'identifiant de l'école sélectionnée auquel on ajoute 300. Par exemple, la première école de remplacement de l'école initialement échantillonnée 056 aura l'identifiant 356.
 - Pour la seconde école de remplacement (R2), l'identifiant de l'école initialement échantillonnée est augmenté de 600. Dans l'exemple précédent, la seconde école de remplacement de l'école 056 recevra l'identifiant 656.

Par exemple, l'école dont l'identifiant PASEC est 012003 est la troisième école échantillonnée (**003**) dans la première strate (**01**) de l'échantillon principal et a été tirée pour les deux niveaux d'enseignement (**2**) concernés. La première école de remplacement aura pour identifiant 012303 et la seconde école, l'identifiant 012603. L'école dont l'identifiant est 031015 est la quinzième école échantillonnée (**015**) dans la troisième strate (**03**) et a été tirée uniquement pour l'évaluation de la classe de fin primaire (5^e / 6^e année).

3.2.7.6 *Finalisation de la liste des écoles échantillonnées*

Une fois les procédures d'échantillonnage terminées, seule la liste des écoles qui doivent participer à l'évaluation et de leurs écoles de remplacement respectives est transmise à l'équipe nationale PASEC de chaque pays pour organiser la collecte des données. Cette liste contient toutes les informations de la base de sondage auxquelles s'ajoutent les informations d'échantillonnage (ID_PASEC, STATUT6, STATUT2, ...). Les premières servent à localiser les écoles et les secondes à gérer de façon efficace la collecte des données sur le terrain.

3.3 L'échantillonnage des classes et des élèves dans les pays participants

L'échantillonnage des classes et des élèves est réalisé le premier jour de l'enquête par les administrateurs de tests envoyés dans les écoles échantillonnées. La procédure d'échantillonnage des classes et des élèves consiste à effectuer un tirage aléatoire simple à niveau.

Dans chaque école échantillonnée, une seule classe est sélectionnée pour le test de fin du primaire (5^e/6^e année). Il en est de même pour le début du primaire (2^e année). Pour le tirage, l'administrateur doit noter le nombre de classes du niveau d'enseignement cible organisées dans cette école et procéder au tirage aléatoire d'une seule classe. Par exemple, s'il y a, dans une école échantillonnée, 3 classes de fin primaire, 6^e année, identifiées 6A, 6B et 6C, l'administrateur doit tirer au sort une classe parmi celles-ci.

Si l'école compte une seule classe de fin primaire (5^e/6^e année) ou de début primaire (2^e année), celle-ci est automatiquement sélectionnée et c'est dans cette classe que les élèves seront sélectionnés aléatoirement pour passer les tests.

Après la sélection de la classe, l'échantillon d'élèves est tiré parmi les élèves du niveau à enquêter inscrits dans la classe. Si la classe comprend 25²³ élèves ou moins en fin de primaire (5^e/6^e année), tous les élèves de la classe sont automatiquement sélectionnés. Parmi eux, seuls les élèves présents passent les tests. Les cas d'élèves échantillonnés mais absents sont assimilés à des non-réponses et leur nombre est répertorié dans la fiche de suivi des élèves.

Si la classe sélectionnée comprend plus de 25 élèves, un échantillon de 25 élèves est constitué par tirage aléatoire simple parmi l'ensemble des élèves du niveau à enquêter inscrits dans cette classe.

3.4 Attribution des cahiers aux élèves de la 6^e année et aux enseignants des écoles

Quatre livrets (A, B, C et D) sont prévus pour les élèves de fin primaire (5^e/6^e année) et quatre livrets (A, B, C et D) pour les enseignants des écoles enquêtées. Après l'échantillonnage des élèves, les livrets élèves sont distribués de façon aléatoire à tous les élèves sélectionnés pour le test. De même, les livrets enseignants sont distribués de façon aléatoire à tous les enseignants des écoles sélectionnées pour le test.

²³ La taille maximale d'élèves à évaluer en fin primaire (5^e/6^e année) est de 25 élèves par classe. Elle est de 16 élèves en début primaire (2^e année)

4. COLLECTE ET SAISIE DES DONNÉES DANS LES PAYS

Les équipes nationales PASEC ont été sensibilisées aux procédures standardisées lors de la mission d'appui-supervision des conseillers techniques. À chaque phase de l'enquête, les pays, par l'intermédiaire des équipes nationales PASEC, ont été guidés à suivre les procédures indiquées par le PASEC pour assurer la comparabilité des résultats. Des points de contrôle ont été effectués tout au long du déroulement de l'enquête pour s'assurer de la conformité du processus. La mise en œuvre de la collecte et de l'encodage des données a été réalisée au cours de l'année 2019 dans les pays participants à l'évaluation internationale PASEC2019 conformément aux procédures du programme. Toutes les opérations ont été réalisées sous la supervision des équipes nationales PASEC. Dans chaque pays, les administrateurs de tests et les agents de saisie ont été recrutés par l'équipe nationale et formés à cet effet. Les formations ont été conduites par l'équipe nationale appuyées et supervisées par les conseillers techniques du PASEC.

Avant la mise en œuvre de la collecte des données, les équipes nationales PASEC ont été formées aux procédures PASEC de collecte des données, d'archivage, de sécurisation des documents et données. Après cette collecte, les équipes nationales sont aussi formées aux procédures de saisie des données.

4.1 Collecte des données

Dans cette section, les principales étapes de la collecte des données de l'évaluation PASEC2019, de la préparation de l'enquête jusqu'à la collecte des données dans les écoles par les administrateurs de tests seront présentées.

L'étape de collecte de données dans les écoles a été réalisée par des administrateurs de tests recrutés et formés par les équipes nationales PASEC sur des contenus communs à tous les pays. Afin d'assurer le caractère comparable pour tous les pays, la formation des administrateurs de tests a été faite sous la supervision des conseillers techniques du PASEC.

Durant les opérations sur le terrain, les équipes nationales PASEC se sont assurées que les administrateurs ont respecté les différentes étapes de l'administration des tests dans les écoles et ont appliqué les standards méthodologiques fixés dans les procédures PASEC.

Le PASEC a appuyé les équipes nationales en présentiel et à distance dans les pays lors de ces différentes phases.

4.1.1 Préparation de la collecte des données

4.1.1.1 Préparation logistique

La phase de collecte des données dans les écoles s'est opérée entre avril et mai de l'année scolaire 2018-2019 sur une période maximale de deux mois.

Les pays ont planifié la période de l'enquête dans les écoles en fonction du calendrier scolaire de leur système éducatif, de l'organisation des examens de fin d'année et des contraintes climatiques ou autres.

Les écoles sélectionnées pour l'enquête par le PASEC ont été réparties par l'équipe nationale entre les administrateurs de tests de manière à regrouper les écoles d'une même zone géographique.

Chaque école a été impartie à deux administrateurs de tests responsables de collecter respectivement les données de début ou de fin de scolarité primaire et du test des enseignants. En somme, chaque administrateur a enquêté 3 écoles sur une période de 3 semaines, soit une école par semaine. Par ailleurs, à chaque école à enquêter sont assignées deux fiches, une fiche de suivi pour l'école et une autre pour les élèves. Celles-ci permettent d'en identifier les caractéristiques et de collecter des informations sur ces entités.

4.1.1.2 Fiches de suivi des écoles et des élèves

Ces formulaires sont préparés par l'équipe nationale PASEC avant les opérations dans les écoles et remplis par les administrateurs de tests au fur et à mesure que l'enquête est exécutée dans les écoles pour enregistrer des informations sur les élèves. Cette phase est contrôlée à distance par le PASEC.

4.1.1.3 Instruments et matériel

Les instruments PASEC sont la propriété unique du PASEC et de la CONFEMEN et ne peuvent être reproduits ou copiés hors du cadre de l'évaluation internationale PASEC. Toute personne en contact avec des épreuves et procédures PASEC s'engage à respecter la confidentialité en signant un engagement écrit.

L'impression des instruments d'enquête est réalisée dans chaque pays. Le PASEC a transmis aux équipes nationales une version électronique et des prototypes des instruments de mesure en version papier. Les versions papier des outils permettent de présenter les normes d'impression aux sociétés d'imprimerie chargées de la reproduction des outils conformément aux normes PASEC. L'imprimeur a été sélectionné à l'issue d'un appel d'offres dont les critères de coût, de rapidité et de qualité prévalent.

Chaque lot de documents imprimés a été vérifié lors de la livraison, en particulier en ce qui concerne la quantité et la qualité de l'impression des cahiers élèves. Une mauvaise qualité d'impression (illisibilité, pages inversées, etc.) peut affecter le fonctionnement des items et la qualité des résultats. Des copies supplémentaires ont par conséquent été imprimées afin de remplacer tout document défaillant.

Les documents sont stockés dans un local sécurisé permettant d'accueillir tous les documents, de les classer et de préparer les colis. Il est attendu que la salle de saisie des données soit adjacente à la salle de stockage des données pour faciliter cette opération. Les instruments sont détruits par l'équipe nationale PASEC un an après la publication des résultats.

Le PASEC a mis à disposition des pays des minuteurs standards pour la passation du test de début de scolarité primaire et des autocollants standards pour sceller les tests. Du matériel de papeterie a été nécessaire pour la préparation des colis (sacs de transport, gommes et crayons, ardoises et craies, etc.).

4.1.1.4 Préparation des colis

Les équipes nationales PASEC ont procédé au scellage des tests et des questionnaires avec des autocollants lors de la préparation des colis. Ce procédé a permis d'assurer la sécurité des items avant l'administration des tests dans les écoles jusqu'au retour des colis au centre national. Il assure aussi que les élèves et les enseignants qui répondent aux questions dans une partie du cahier ne peuvent répondre à celles d'une autre partie.

Les documents et matériels nécessaires à la collecte des données dans les écoles ont été consignés (cahier Élèves, cahier Enseignant, questionnaires Directeurs, cahier d'administrateur de tests, crayons, gommes, minuteur pour le début de primaire dans des sacs auxquels sont agrafés une fiche avec le nom de l'école et le numéro de l'administrateur. Cette phase a été contrôlée par le PASEC lors d'un appui dans les pays.

4.1.2 Formation des administrateurs de tests

4.1.2.1 Recrutement des administrateurs de tests

Les administrateurs de tests sont recrutés dans les pays par l'équipe nationale PASEC et formés dans la capitale avant leur départ dans les écoles.

Les équipes nationales PASEC ont suivi les normes de recrutement fixées par le PASEC pour recruter les administrateurs de tests : titulaires d'un diplôme dans le domaine de l'enseignement, enseignants et/ou directeurs déchargés de leur poste ou à la retraite. Le PASEC a recommandé de recruter des administrateurs expérimentés, dynamiques et motivés et de ne pas recruter du personnel éducatif en fonction dans les écoles pour des raisons éthiques.

Le processus de recrutement et de formation a été déterminant pour l'évaluation en début de scolarité primaire dans la mesure où la fidélité et la fiabilité des réponses des élèves reposent essentiellement sur la qualité des administrateurs de tests responsables de donner les consignes de passation puis de retranscrire les réponses des élèves. Les administrateurs de tests devraient être disponibles pendant toute la durée de l'enquête et de la formation des administrateurs, soit près d'un mois.

Chaque administrateur participant à la formation s'est par ailleurs engagé à respecter les clauses de confidentialité des tests et des procédures PASEC.

4.1.2.2 Formation théorique et pratique des administrateurs de tests

Les équipes nationales PASEC ont été sensibilisées aux procédures de passation lors des regroupements internationaux et des appuis dans les pays offerts par le PASEC.

Tous les administrateurs de tests recrutés pour l'enquête ont suivi une formation théorique et pratique sur les procédures de passation des tests dans les écoles. Ces procédures ont été rassemblées dans un cahier d'administrateur qui retrace toutes les étapes à respecter pour collecter les données conformément aux standards du PASEC. Les tests étant standardisés, tous

les élèves doivent être soumis aux mêmes conditions de passation et l'administrateur ne doit donc pas s'écarter des consignes du PASEC.

Les formations ont été dispensées sur 3 jours pour les administrateurs de tests de fin de scolarité primaire et enseignants et sur 4 jours pour les administrateurs de début de primaire puisque ce dernier est en passation individuelle avec les élèves. Cette activité a été assurée par l'équipe nationale PASEC avec l'appui du PASEC, le cas échéant.

- **Formation théorique**

La formation théorique des administrateurs de tests a été cruciale pour le bon déroulement de la collecte des données. Tous les documents, hormis les tests de fin du primaire et les enquêtes enseignant (passation en autonomie complète par les élèves et les enseignants), ont été présentés en profondeur aux administrateurs pour partager les bonnes pratiques et les expériences antérieures. Des simulations ont permis d'illustrer les procédures.

En début de scolarité primaire administrer un test revient à soumettre une suite d'items à un élève en passant d'un exercice à un autre dans l'ordre du cahier Élèves. Il est important de bien former l'administrateur à l'utilisation combinée du cahier Élèves et du support Élèves.

En fin du primaire et chez les enseignants, l'administration des tests consiste essentiellement à créer les conditions optimales pour que les élèves et les enseignants remplissent les cahiers de tests dans de bonnes conditions, en ayant compris et assimilé toutes les consignes.

Durant cette formation, les questionnaires Élèves, Maîtres et Directeurs sont également présentés.

- **Formation pratique et simulation dans une école**

Pendant la mise à l'essai, la phase pratique a permis aux administrateurs de compléter leur formation théorique en procédant à la simulation d'une enquête dans une école témoin. Lors de cette activité, plusieurs administrateurs de tests ont été sélectionnés pour prendre part à une partie de l'enquête : présentation de l'enquête, tirage de la classe, tirage des élèves, passation d'un test, etc.

Durant cette journée, tous les administrateurs ont observé la simulation et relevé les difficultés et bonnes pratiques à chacune des étapes de l'enquête dans l'école.

Une réunion-bilan a été organisée avec tous les administrateurs de tests pour échanger sur les pratiques et corriger les difficultés observées.

La simulation dans une école n'est pas menée avec les grands groupes lors de l'agenda final en 2019 en raison de la complexité de coaching et de la supervision de ces groupes dans l'école. En substitution de cette immersion en situation réelle, les simulations ont été accrues à travers des jeux de rôle en petits groupes et/ou en plénière. La formation à l'enquête finale a suivi le même processus de formation des administrateurs en parallèle.

4.1.2.3 Sélection des administrateurs de tests

Tout au long de la formation, les administrateurs ont été encadrés par l'équipe nationale PASEC. En fin de formation, les administrateurs ont été évalués sur leur connaissance des procédures de passation. Les administrateurs ont été notés à la fin de la formation et ceux ayant les notes les plus basses n'ont pas été sélectionnés.

À cette étape, les administrateurs retenus ont reçu un identifiant d'administrateur, un colis et les fiches de suivi des écoles pour cibler les écoles à enquêter.

Chaque membre de l'équipe nationale PASEC a eu la responsabilité d'assurer la supervision d'une partie des administrateurs lors des opérations terrain selon une couverture géographique préétablie.

4.1.3 Collecte des données

4.1.3.1 Collecte des données dans les écoles

Les procédures de collecte des données dans les écoles sont consignées dans le cahier d'administrateur de tests communs à tous les pays. Ce document indique les standards à respecter et détaille les tâches et points de contrôle à effectuer par les administrateurs de tests au cours de la collecte des données. La collecte des données dans chaque école suit le programme suivant :

Tableau : Programme de la collecte des données dans les écoles

Journée	Jour 1 :	Jour2 :	Jour 3 :	Jour 4 :
Enquête	Présentation de l'enquête, tirage des classes et des élèves de 2A et 6A Administration des tests et questionnaires aux élèves selon le programme de passation retenu dans l'école	Administration des tests et questionnaires aux élèves selon le programme de passation retenu dans l'école	Administration des tests et questionnaires aux élèves selon le programme de passation retenu dans l'école	Administration des tests et questionnaires uniquement aux élèves de début de scolarité primaire selon le programme de passation retenu dans l'école Administration de tout le programme du questionnaire enseignant Vérification des données récoltées et des instruments

4.1.3.2 Préparation de l'enquête et tirage des classes et élèves

Les jours de collecte dans les écoles, les administrateurs de tests ont vérifié tout leur matériel d'évaluation et rapporté à leur superviseur tout problème d'accessibilité aux écoles. Les administrateurs ont été informés par le PASEC et l'équipe nationale que l'école à enquêter ne peut être remplacée qu'en cas de force majeure ou de refus définitif et non pas par convenance. L'administrateur n'a procédé au remplacement d'une école qu'avec l'accord du superviseur.

Chaque administrateur a également suivi l'ordre des écoles indiqué dans les fiches de suivi de son cahier d'administrateur et est arrivé tôt le matin afin de disposer de suffisamment de temps pour présenter sa mission et les objectifs de l'enquête au directeur.

Les administrateurs ont ensuite procédé au tirage aléatoire de la classe, le cas échéant, parmi l'ensemble des classes du niveau enquêté présentes dans l'école, puis au tirage des élèves inscrits dans la classe.

4.1.3.3 Passation des tests de début de scolarité primaire dans les écoles

Les écoles sont responsables d'offrir à leurs élèves des conditions satisfaisantes de passation des tests. Les tests ont été administrés individuellement à chacun des 16 élèves de début de scolarité primaire sélectionnés ; pour ce faire, les administrateurs avaient pour consigne de livrer les instructions et les questions au mot près afin de garantir la standardisation des tests dans tous les pays. Les tests se sont déroulés sur quatre matinées, soit deux matinées par discipline, et l'ordre de passation des tests a été préalablement déterminé de façon aléatoire pour chaque école. Le temps de passation des tests et questionnaires a été approximativement de 30 minutes par test et par élève chaque jour.

Les administrateurs avaient pour consigne de commencer par mettre les élèves en confiance. En effet, le mode de passation des tests ne correspond pas, ou très peu, aux pratiques de classe et d'évaluation habituelles à l'école primaire. Les administrateurs ont ensuite complété la fiche de suivi des élèves et administré le test. Les procédures de passation ont été suivies scrupuleusement et aucune appréciation orale des réponses des élèves n'a été donnée. Seuls les encouragements ont été autorisés.

Les administrateurs de tests disposent pour chaque élève d'un cahier Élèves nominatif qui contient les consignes de passation et les tableaux de correction dans lesquels ils corrigent directement les réponses des élèves. Ils mettent à la disposition des élèves, pour la majorité des exercices, un support Élèves contenant des images, des grilles de lettres et de mots et des textes que les élèves doivent parcourir et lire pour répondre aux différents exercices. En mathématiques, les élèves disposent également d'une ardoise et de craies les aidant à effectuer les opérations et à résoudre les problèmes.

La police cursive utilisée dans le support *Élèves vise*, autant que possible, à reproduire l'écriture des maîtres dans les salles de classe.

4.1.3.4 Passation des tests de fin de scolarité primaire dans les écoles

Les administrateurs disposaient d'une salle de classe équipée de bancs et de bureaux pour y faire passer le test. Le temps de passation des tests et questionnaires de fin du primaire (passation collective) était de 2 heures maximum par matinée, le premier jour est consacré au tirage de la classe et des élèves participants et questionnaires, les deuxièmes et troisièmes jours sont dédiés aux tests de compréhension de l'écrit et mathématiques selon l'ordre indiqué établi par le PASEC. Le test dans chaque discipline est organisé en 2 séquences d'une heure séparées par une pause de 10 minutes. Le quatrième jour est réservé à l'enquête sur les enseignants de l'école. L'évaluation s'étendait sur quatre jours et l'ordre de passation des tests en fonction de la discipline (lecture ou mathématiques) était indiqué au préalable dans la fiche de suivi de l'école. L'administrateur ne pouvait changer cet ordre.

En préparation de la passation des tests, l'administrateur a réparti les livrets de la façon suivante :

- Livret A : Élèves 1 à 6 sélectionnés
- Livret B : Élèves 7 à 13 sélectionnés
- Livret C : Élèves 14 à 19 sélectionnés
- Livret D : Élèves 20 à 25 sélectionnés

La première partie du test de langue ou de mathématiques a été descellée. Avant de la soumettre aux élèves, les administrateurs se sont assurés que ces derniers ont bien compris les consignes et les modalités de réponse aux questions. Pour cela, les administrateurs ont procédé à quelques exemples standards à voix haute et répondu aux questions des élèves. En aucun cas les administrateurs ne devaient apporter de l'aide aux élèves pour trouver les réponses aux questions, mais jouaient plutôt un rôle de relance auprès des élèves qui étaient bloqués sur une question. Durant le test, les administrateurs se sont aussi assurés que les élèves ne trichaient pas.

Une fois la première heure écoulée, les élèves bénéficiaient d'une pause de 10 minutes permettant aux administrateurs de sceller le test effectué et de desceller la seconde partie. Le même processus de passation a été respecté pour l'administration de la seconde partie. Les administrateurs ont répété la même démarche le lendemain avec la matière non évaluée.

4.1.3.5 Passation des questionnaires contextuels dans les écoles

La passation du questionnaire directeur-école a eu lieu tout au long de l'évaluation PASEC. Les questionnaires Élèves 6A ont été administrés le 1^{er} jour de l'enquête ; les autres questionnaires (Elèves 2A et maîtres ont été administrés simultanément avec les tests).

Ces questionnaires ont pour but de donner des indications contextuelles.

En début du primaire, l'administration du questionnaire Élèves s'est faite en passation individuelle avec l'administrateur, qui a rempli le questionnaire en fonction des réponses de chaque élève.

Cette passation pouvait s'effectuer dans la langue de socialisation de l'élève pour faciliter sa compréhension, si l'administrateur la parle aussi.

En fin de scolarité primaire, la passation des questionnaires a été collective : les administrateurs lisaient les questions et les modalités de réponse et s'assuraient que tous les élèves remplissaient le questionnaire en même temps.

L'administration des questionnaires Maîtres et Directeurs a suivi un seul et même processus. Les maîtres des classes évaluées ont rempli le questionnaire. Le remplissage du questionnaire maîtres et Directeurs relevait de l'administrateur de fin de scolarité primaire. Les maîtres et le directeur répondaient directement aux questionnaires. Les administrateurs pouvaient, le cas échéant, répondre à leurs interrogations et vérifiaient que toutes les questions comportaient une réponse. La seconde partie du questionnaire directeur était remplie avec l'administrateur, qui vérifiait les réponses des directeurs selon les documents et équipements disponibles dans la classe et l'école.

Une fois les tests et les questionnaires répondus, les administrateurs devaient vérifier que les questionnaires étaient bien remplis et que les cahiers Élèves étaient scellés. Ils s'assuraient également de ne pas laisser de documents dans l'école enquêtée. Le processus d'évaluation s'est ensuite répété dans les autres écoles en fonction du programme fixé à l'attention des administrateurs.

4.1.3.6 *Retour des instruments du terrain*

Tous les outils d'enquête ont été vérifiés par l'équipe nationale PASEC avec les administrateurs lors du retour des outils. Des sanctions ont été appliquées si des non-réponses sur plus de 5 % des questions ont été constatées.

4.2 Rangement et sécurisation des instruments de collecte revenus du terrain

Dans chaque pays participant à l'évaluation internationale PASEC2019, les instruments revenus du terrain ont été rangés par ordre croissant d'identifiant PASEC (écoles) et d'identifiant élèves ou enseignant de façon à former 10 lots par type d'instrument. Ce rangement a grandement contribué à la résolution de la plupart des problèmes rencontrés lors de la saisie et du nettoyage des données.

Pour un pays ayant échantillonné 180 écoles de 6^e année et 90 écoles de 2^e année, l'ensemble des instruments revenus du terrain est composé²⁴ de :

- un minimum de 180 fiches de suivi des écoles pour la 6^e année. Les fiches de suivi des écoles échantillonnées non enquêtées de même que celles des écoles remplaçantes doivent figurer dans le lot ;
- un minimum de 90 fiches de suivi des écoles pour la 2^e année. Les fiches de suivi des écoles échantillonnées non enquêtées de même que celles des écoles remplaçantes doivent figurer dans le lot ;
- un maximum de 180 fiches de suivi des élèves pour la 6^e année. Pour les écoles non enquêtées, les fiches de suivi des élèves ne sont pas prises en compte ;
- un maximum de 90 fiches de suivi des élèves pour la 2^e année. Pour les écoles non enquêtées, les fiches de suivi des élèves ne sont pas prises en compte ;
- un maximum de 900 livrets A et questionnaires Élèves pour la 6^e année ;
- un maximum de 900 livrets B et questionnaires Élèves pour la 6^e année ;
- un maximum de 900 livrets C et questionnaires Élèves pour la 6^e année ;
- un maximum de 900 livrets D et questionnaires Élèves pour la 6^e année ;
- un maximum de 900 livrets et questionnaires Élèves pour la 2^e année ;
- environ de 270 livrets A et questionnaires Enseignants ;
- environ de 270 livrets B et questionnaires Enseignants ;
- environ de 270 livrets C et questionnaires Enseignants ;
- environ de 270 livrets D et questionnaires Enseignants ;
- un maximum de 180 questionnaires Directeurs.

Ainsi, dans chaque pays, il y a eu environ 10 lots de chaque type d'instrument.

Le tableau suivant illustre la constitution préalable des lots d'écoles et le réajustement après la collecte pour chaque type d'instrument.

²⁴ Les quantités mentionnées sont établies sur la base d'un échantillon standard de 180 écoles de 6^e année et de 90 écoles de 2^e année du primaire. Ils ne tiennent pas compte des possibles augmentations ou diminutions (pertes) de la taille des échantillons.

Tableau 4.2 : Exemple de constitution des 10 lots

École	ID_PASEC	Lot
1	012001	1
2	011002	1
3	012003	1
...
19	012019	2
...
37	022001	3
...
57	021021	4
58	022022	4
...

Dans ce tableau, l'école n° 1 de la liste est celle dont l'identifiant est 012001. Cette école a été classée avant l'école n° 2, dont l'identifiant est 011002.

Si, lors de la collecte, l'école n° 1 est remplacée par sa première école de remplacement, dont l'identifiant PASEC est 012301, cette école de remplacement sera aussi classée avant l'école n° 2 dont l'identifiant est 011002.

Ce travail de classement doit donc être fait à l'intérieur des lots. Du lot 1 au lot 10, le classement des écoles suit l'ordre de numérotation des strates indiqué par les deux premiers chiffres de l'identifiant PASEC alors que la numérotation des écoles à l'intérieur de la strate est représentée par les trois derniers chiffres.

Par exemple, la première école de la deuxième strate dont l'identifiant PASEC est 021001 viendra bien après la première école de la première strate dont l'identifiant PASEC est 012001.

Les instruments de collecte ainsi que les ordinateurs utilisés pour la saisie des données ont été placés dans un local sécurisé avec accès réglementé pour assurer la sécurité et la confidentialité des données collectées. Toutes les personnes ayant eu un contact quelconque avec le matériel sécurisé ont signé une clause de confidentialité.

4.3 Saisie des données

La codification des données a précédé la saisie des données. Cette tâche a été confiée aux agents de codification qui se sont basés sur les grands principes de l'encodage des données en parcourant les informations collectées dans un instrument :

- D'abord de gauche à droite puis de haut en bas ;
- Les chiffres suivent la même logique de 1 à n en incrémentant chaque fois de 1 (de gauche à droite et de haut en bas) ;
- Il existe différents types de données manquantes :
 - Le répondant (élève, maître ou directeur) devait répondre mais n'a pas répondu. Si le masque de saisie autorise la saisie d'un seul chiffre, alors le code 9 est utilisé pour ce type de donnée manquante. Si le masque de saisie autorise deux chiffres, alors le code 99 sera utilisé, etc. ;
 - Le répondant ne devait pas répondre à la question, comme dans le cas des questions emboîtées (« si oui, répondre à la question suivante, si non, passer à la question x ») ou en raison du plan d'évaluation incomplet pour les tests de mathématiques ou de compréhension de l'écrit. Le code 7, 97 ou 997 et ainsi de suite est attribué en fonction du nombre de chiffres utilisés dans le masque de saisie ;
 - Le répondant a fourni une réponse invalide, soit en cochant plus d'une case alors qu'il devait en sélectionner une seule, soit en répondant n'importe quoi à une question ouverte. Le code 6, 96 ou 996 et ainsi de suite est attribué en fonction du nombre de chiffres utilisés dans le masque de saisie.

L'application de ces grands principes a conduit aux codifications suivantes :

4.3.1 Codification des instruments des tests de 6^e année

Les réponses²⁵ aux items de 6^e année identifiée par les cases cochées prennent les valeurs 1, 2, 3, 4, 6 ou 9 (6=réponse invalide et 9=non-réponse).

4.3.2 Codification des instruments des tests de 2^e année

Les réponses aux items de 2^e année (autre les grilles de lettres, de mots ou de nombres) sont entourées par les administrateurs et prennent les valeurs 1 (correct), 2 (incorrect), 6 (réponse invalide) et 9 (non réponse) : le code 6 en 2^e année indique que l'administrateur a entouré plusieurs réponses et le 9, que l'administrateur n'a rien inscrit.

En lecture, les réponses aux grilles de lettres ou de mots en 2^e année sont codées en fonction des réponses sur la grille.

En mathématiques, les réponses aux grilles de nombres en 2^e année sont codées en fonction des réponses sur la grille. La première tâche des agents de codification consiste à identifier l'un des deux cas de figure possibles avant de commencer la codification :

²⁵ La recodification des items non atteints et des données manquantes par design a été effectuée lors du nettoyage des données.

1^{er} cas : le code correspond uniquement au nombre entouré. Le code doit être compris entre 1 et 100 ; on inscrit 996 si la réponse est invalide (des nombres sont barrés ou il y a des traits alors qu'on n'en demande pas, ou bien plusieurs nombres sont entourés) ou 999 en cas de non-réponse (non-correction de l'administrateur) ;

2^e cas : le code correspond à la logique suivante :

- 1=nombre non barré (nombre atteint et bien lu) ;
- 2=nombre barré (nombre atteint et mal lu) ;
- 6=réponse incohérente : quand un ou des nombres sont barrés et que la case « l'élève a lu correctement tous les nombres » est cochée.
- 9=non-réponse : par exemple, la correction de l'administrateur n'est pas faite, c'est-à-dire qu'aucun nombre n'est barré et que la case « l'élève a lu correctement tous les nombres » n'est pas cochée. Attention : la case « l'élève a lu correctement tous les nombres » à cocher n'est pas codée dans le masque de saisie, elle est là uniquement pour savoir si l'exercice a été fait ou non par l'élève.

4.3.3 Codifications des instruments des questionnaires contextuels

Les modalités sont les suivantes :

- Question à choix multiple à laquelle une seule réponse est attendue : 1=1^{re} réponse, 2=2^e réponse, ..., $n=n^e$ réponse, 6, 96 ou 996 si réponse invalide ou 9, 99 ou 999 si non-réponse, selon le nombre de chiffres autorisé ;
- Question à choix multiple où plusieurs réponses sont possibles : 1=case cochée et 2=case non cochée ;
- Réponse si condition : 1=oui, 2=non, 6=réponse invalide (incohérence : plusieurs réponses entourées ou la personne a coché « non » mais répondu à la question d'après) ou 9=non-réponse ;
- Ne pas oublier les codes où on doit reporter le nombre donné par le répondant, tel que l'âge, l'expérience, le revenu, etc.

Cette codification a été la même pour les dix pays à l'exception des questions dans les questionnaires Maîtres et Directeurs relatives à la classe la plus élevée atteinte par le maître ou le directeur, au diplôme le plus élevé possédé et à leur statut, où le nombre de modalités de réponse pouvait varier d'un pays à un autre.

Pour chaque pays participant, la procédure de saisie des données a consisté en la réalisation de plusieurs activités, à savoir :

- le développement des masques de saisie ;
- le recrutement et la formation des agents de saisie ;
- l'organisation générale de la saisie ;
- le contrôle de qualité ;
- la sauvegarde des données ;
- la communication entre l'équipe nationale et le PASEC.

4.3.4 Développement des masques de saisie

Le PASEC a développé les masques de saisie par type d'instrument à travers un logiciel de saisie. Ces masques comportent des contrôles n'autorisant, pour chacune des cellules de saisie, que des données respectant certaines conditions (de format, de valeur...).

Par exemple, il n'est pas possible d'écrire du texte dans une cellule qui attend une donnée numérique ou de saisir la valeur « 21 » dans une cellule pour laquelle la valeur maximale est fixée à 20. Ces masques ont été mis à la disposition des équipes nationales PASEC. De nombreux outils ont également été développés et partagés avec les équipes nationales PASEC pour faciliter le recrutement des agents de saisie, le contrôle de la qualité de la saisie et l'organisation des fichiers de saisie.

Afin de veiller au respect strict des procédures de saisie et des délais pour la mise à disposition des données, cette phase a été contrôlée par le PASEC lors d'un appui dans chacun des dix-neuf pays participants. Ceci a permis de contribuer à la formation des agents de saisie et de renforcer les capacités des équipes nationales dans la supervision et le contrôle des activités de saisie.

Les vérifications menées par le PASEC dans chacun des pays peuvent se résumer de la manière suivante :

- Le contrôle du respect des procédures de stockage des instruments après la collecte de données, conformément aux procédures de gestion de données du PASEC, et l'aide à la résolution des problèmes rencontrés lors de cette première étape ;
- L'identification des entraves à la mise en place des conditions idéales de codification et de saisie (disponibilité de la salle et du minimum de dix ordinateurs) et, au cas où ces conditions ne sont pas remplies, la proposition d'une réorganisation qui permettrait d'obtenir des données fiables et dans des délais raisonnables ;
- Le renforcement des capacités de l'équipe nationale dans l'élaboration et la manipulation des masques de saisie sur EpiData, en complément des présentations et applications réalisées lors du deuxième regroupement des équipes nationales ;
- L'opérationnalisation de la mise en œuvre des procédures de saisie (supervision et contrôle de qualité) pour les masques de saisie élaborés par le PASEC (contrôle au niveau de la double saisie, correction des erreurs, fusion des données à partir de plusieurs masques, etc.) ;
- Le lancement avec l'équipe nationale PASEC des opérations de saisie.

4.3.5 Recrutement et formation des agents de codification et de saisie

Conformément aux procédures de codification et de saisie des données, des agents de codification et de saisie ont été recrutés dans chaque pays. Ces agents de codification et de saisie ont dû se soumettre respectivement à un test de codification des données d'un questionnaire et à un test d'utilisation d'un masque de saisie réalisé avec le logiciel EpiData fourni par le PASEC. Le test de saisie d'une dizaine de minutes consistait à saisir les données d'un questionnaire « test » contenant un identifiant, quelques données textuelles et quelques données numériques.

Le postulant pour la codification ou la saisie est considéré comme compétent s'il parvient à codifier ou saisir l'ensemble des données d'un questionnaire sans commettre d'erreur et dans le temps accordé.

Une fois recrutés, les agents de codification et les opérateurs de saisie ont été formés respectivement à la codification des données et à l'utilisation des masques de saisie à partir d'un manuel de formation élaboré par le PASEC.

4.3.6 Saisie des données et contrôle de qualité

La saisie des données s'est organisée en trois phases distinctes dans chaque pays et a duré entre 30 et 45 jours.

La supervision et le contrôle de la saisie dans chaque pays ont été assurés par deux membres de l'équipe nationale (le statisticien et l'informaticien). Une phase quotidienne de double saisie partielle a permis un suivi rapproché de la qualité du travail de saisie.

Chaque journée de travail est clôturée par une phase d'une heure de contrôle de qualité qui consiste à ressaisir une partie des données enregistrées au cours de la journée. Ainsi, 10 % des fiches de suivi (des écoles et des élèves) et 15 % des livrets et questionnaires ont été choisis aléatoirement (par un code informatique mis à la disposition des équipes nationales par le PASEC) et soumis à une double saisie.

Concrètement, le contrôle de qualité a été organisé de la manière suivante en fonction des différentes phases : chaque opérateur a effectué la double saisie d'instruments initialement saisis par un autre opérateur. Le jour suivant, le superviseur, tout en continuant son rôle de supervision, a analysé la conciliation automatique des données saisies le jour précédent. Lorsqu'une incohérence entre la saisie initiale et la double saisie était identifiée par le logiciel, le superviseur vérifiait la donnée exacte dans l'instrument. Après vérification et correction des incohérences, tous les instruments qui ont été soumis à la double saisie ont été directement reclassés dans les lots correspondants.

La saisie initiale est l'enregistrement qui doit être validé et corrigé au besoin, la double saisie ne servant qu'à sa vérification. Au terme de la vérification de la double saisie d'un opérateur, le superviseur a calculé le taux d'erreurs par rapport à l'ensemble des documents sélectionnés (nombre d'erreurs détectées au fur à mesure par rapport au nombre de variables contenues dans un instrument multiplié par le nombre d'instruments sélectionnés aléatoirement)²⁶.

Si le taux d'erreurs dépassait 5 %, le superviseur supprimait toutes les données de la saisie initiale et de la double saisie (pour ne pas conserver deux enregistrements pour une même école) et l'opérateur reprenait toute la saisie le lendemain.

Afin d'éviter des doublons dans un fichier, la double saisie a été réalisée dans un fichier différent de celui de la saisie initiale. Cette opération nécessitait que le superviseur tienne un registre précis des différents lots traités par chacun des opérateurs afin de déterminer, avec précision et sans commettre d'erreur, les lots qui devaient être saisis de nouveau. Les instruments ressaisis à la suite d'un taux d'erreurs d'au moins 5 % étaient à nouveau contrôlés. Si ce taux ne dépassait pas les 5 %, le superviseur corrigeait les erreurs détectées dans les données de la saisie initiale et supprimait les données de la double saisie.

Les erreurs commises par chaque opérateur dans la saisie initiale ont été comptabilisées afin de calculer leur taux d'erreurs quotidien.

À la fin de la journée, à partir de la deuxième journée de saisie, le superviseur a donné à chacun des opérateurs une rétroaction sur la qualité de son travail de la veille et sur la nature des erreurs commises par chacun. Les sanctions éventuelles ont été appliquées comme décrit plus haut.

Dans tous les pays, aucun opérateur n'a excédé la barre des 10 % d'erreurs. Bien souvent, les taux d'erreurs enregistrés ont été inférieurs à 5 %. Étant donné le faible taux d'erreurs constaté dans les pays, le travail des opérateurs n'a été vérifié qu'un jour sur deux. Les taux d'erreurs enregistrés sont restés faibles.

²⁶ Voir le manuel des procédures du PASEC à la partie relative à la saisie des données.

4.3.7 Sauvegarde des données

Les superviseurs de la saisie des données ont procédé quotidiennement à la récupération des données saisies, par type d'instrument, sur un support externe. Cette opération permettait, en cas de manipulations inappropriées sur une machine, de perte d'ordinateurs ou de tout autre incident, de conserver un maximum de données déjà saisies.

4.3.8 Communication entre les équipes nationales et le PASEC

Pendant toute la période de saisie, le responsable national de chaque équipe PASEC a fourni au PASEC un bilan hebdomadaire de la progression des travaux de saisie des données en se basant sur les comptes rendus des superviseurs. Il a aussi envoyé, à la fin de chaque semaine et pour chaque type d'instrument, la base de données fusionnée et corrigée qui intégrait en un seul fichier le travail de tous les agents de saisie. Cette stratégie a permis au PASEC de suivre efficacement la progression des activités sur le terrain.

5. NETTOYAGE DES DONNÉES

Après la collecte et la saisie des données issues de l'évaluation internationale PASEC2019, les équipes nationales des quatorze pays participants ont chacune transmis au PASEC 12 bases de données, à savoir :

- une base de données des fiches de suivi des écoles pour la 5^e/6^e année ;
- une base de données des fiches de suivi des écoles pour la 2^e année ;
- une base de données des fiches de suivi des élèves pour la 5^e/6^e année ;
- une base de données des fiches de suivi des élèves pour la 2^e année ;
- une base de données des livrets A et des questionnaires Élèves pour la 5^e/6^e année ;
- une base de données des livrets B et des questionnaires Élèves pour la 5^e/6^e année ;
- une base de données des livrets C et des questionnaires Élèves pour la 5^e/6^e année ;
- une base de données des livrets D et des questionnaires Élèves pour la 5^e/6^e année ;
- une base de données des livrets et des questionnaires Élèves pour la 2^e année ;
- une base de données des livrets A et des questionnaires Enseignants ;
- une base de données des livrets B et des questionnaires Enseignants ;
- une base de données des livrets C et des questionnaires Enseignants ;
- une base de données des livrets D et des questionnaires Enseignants ;
- une base de données issue de la saisie des questionnaires Directeurs.

Les données de chacune de ces bases ont été collectées et saisies suivant les procédures recommandées par le manuel de procédures du PASEC. Le nettoyage des données des quatorze pays a été réalisé par le PASEC en collaboration avec les équipes nationales PASEC des pays participants à l'évaluation selon les phases suivantes :

- programmation sous Stata des procédures de nettoyage des données et d'identification des incohérences dans les données ;
- correction des données avec l'appui des équipes nationales ;
- mission d'accélération ou de finalisation du nettoyage des données dans certains pays ;
- finalisation des bases de données : recodage de certains items (non atteints ou manquants par design).

5.1. Programmation sous Stata des procédures de nettoyage des données et identification des incohérences dans les données

Le nettoyage des données de tous les pays participants à l'évaluation groupée du PASEC s'est réalisé conformément aux procédures du PASEC. Le respect scrupuleux de ces procédures a permis de fiabiliser la comparabilité des données internationales. En particulier, toutes les données de tous les pays ont été vérifiées et nettoyées suivant des procédures identiques. Lorsqu'une modification est apportée à un programme pour tenir compte du contexte d'un pays, la même vérification est effectuée pour tous les autres pays.

Comme indiqué précédemment, le PASEC a démarré l'activité de nettoyage des données par une programmation (sous Stata) de ces différentes procédures : un total de 17 programmes informatiques ont été élaborés et ont servi de base pour le nettoyage des données de tous les pays participants à l'évaluation PASEC2019. À ceux-ci se sont ajoutés les programmes spécifiques pour chacun des quatorze pays, et qui ont permis de modifier les bases de données suite aux erreurs de saisie et autres problèmes constatés dans les données.

5.2. Recodage des données avec l'appui des équipes nationales

Pour éviter autant que possible la suppression d'une donnée pour une variable, voire la suppression de l'ensemble des données d'un élève, d'un maître ou d'une école qui conduirait à une diminution de la taille de l'échantillon pouvant biaiser les analyses, le nettoyage des données a été un processus collaboratif entre les équipes nationales et le PASEC. Il a connu plusieurs étapes :

- 1) Vérification et correction des incohérences sur les identifiants au niveau de chaque base de données ;
- 2) Vérification et correction des incohérences entre les différentes bases de données ;
- 3) Vérification et correction des incohérences au niveau interne des bases de données.

5.2.1 Vérification et correction des incohérences au niveau des identifiants dans chaque base de données

Pour chaque pays et chaque base de données, le PASEC a vérifié et corrigé les incohérences au niveau des identifiants des écoles à 6 chiffres et des identifiants des élèves à 2 chiffres (de 01 à 25 pour la 5^e/6^e année et de 01 à 16 pour la 2^e année). Les doublons ont également été vérifiés et corrigés avec la collaboration des équipes nationales.

Les incohérences relevées découlaient d'une mauvaise saisie des identifiants. Dans bien des cas, ces identifiants contenaient des lettres qui n'avaient rien à voir avec le format demandé par le PASEC²⁷.

5.2.2 Vérification et correction des incohérences entre les différentes bases de données

Après la correction des incohérences portant sur les identifiants et les doublons, le PASEC a vérifié, d'une part, la cohérence entre les identifiants des écoles et ceux des élèves dans les différentes bases de données et, d'autre part, entre les déclarations des maîtres et celles des directeurs d'écoles. Cela consistait en :

- la vérification de la concordance des noms et des identifiants des écoles entre les différents instruments et les informations relatives à l'échantillon initial ;
- la vérification de la concordance entre les fiches de suivi des élèves et les livrets (comparaison des identifiants et des noms/prénoms des élèves dans les deux bases, comparaison des nombres d'enregistrements, comparaison des identifiants des livrets...);
- la vérification de la cohérence des déclarations (sur les effectifs) des enseignants et des directeurs dans les questionnaires en lien avec les informations contenues dans les fiches de suivi ;
- la mise en lien de la taille de l'école avec la somme des tailles des classes de 2^e année et de 5^e/6^e année, la première devant être supérieure à la seconde ;
- l'identification des saisies incomplètes : élèves identifiés dans les livrets mais pas dans les fiches de suivi des élèves ou élèves déclarés participants dans les fiches de suivi des élèves mais pas dans les livrets ;

²⁷ Les identifiants des écoles ou des élèves pouvant commencer par le chiffre 0, bien que n'admettant que les chiffres ont un format alphanumérique.

- la vérification de la structure de réponses des élèves (lignes entièrement manquantes alors que les élèves sont déclarés participants).

À l'issue de ces différentes vérifications, les incohérences suivantes ont été identifiées et des mesures de correction ont été entreprises :

- Un nombre non négligeable d'écoles dans certains pays ont été remplacées pour des motifs d'inaccessibilité au moment de la collecte des données. La liste des écoles participantes aux évaluations a été mise à jour et les écoles de remplacement ont été maintenues dans les données en conformité avec les procédures du PASEC en la matière ;
- Certains instruments se sont avérés non saisis par confrontation des données des livrets et des fiches de suivi. Pour tous les pays, les données des instruments existants mais non saisis ont été ajoutées dans les bases de données après vérifications par le PASEC ;
- Dans tous les pays, un nombre important d'élèves ayant réalisé les tests étaient déclarés absents ou exclus dans les fiches de suivi. La vérification de ces déclarations et les corrections ont été faites ;
- Dans tous les pays, des incohérences ont été détectées entre les déclarations des directeurs et celles des enseignants, et ceci aux deux niveaux scolaires évalués (taille des classes enquêtées, effectifs enquêtés, nombre d'élèves pour un niveau donné, etc.). La correction de ces incohérences a été indispensable pour un calcul correct des poids des élèves dans les bases de données finales.

Certaines des incohérences identifiées ont été corrigées par le PASEC et transmises aux équipes nationales pour confirmation. D'autres, par contre, ont été soumises aux équipes nationales pour correction. La validation des corrections proposées par le PASEC ainsi que la correction des autres incohérences ont été réalisées par les équipes nationales après consultation des instruments de collecte et parfois des directeurs des écoles concernées.

5.2.3 Vérification et correction des incohérences au niveau interne des bases de données

Après la vérification et la correction des incohérences entre les bases des données par le PASEC en collaboration avec les équipes nationales, les vérifications ci-dessous ont été réalisées par le PASEC au niveau interne des bases de données :

- identification et correction des données aberrantes/atypiques sur les variables ;
- identification et correction des cas de non-respect des filtres ;
- vérification de la cohérence interne des bases des données ;
- vérification de la concordance dans le codage des items relatifs aux grilles de lettres et de mots dans le livret de 2^e année ;
- etc.

Dans tous les pays, des données atypiques sur certaines variables des questionnaires contextuels ont été identifiées et corrigées en recourant systématiquement aux instruments de collecte ou en appelant les directeurs des écoles concernées (âge de l'élève, nombre de personnes avec qui l'élève vit, âge et ancienneté du maître, nombre de classes tenues par le maître, nombre moyen de leçons de français

ou de mathématiques par semaine, nombre moyen d'élèves absents par mois, nombre effectif d'heures de cours par semaine, nombre de places assises ou de supports pour écrire pour les élèves dans la classe, âge et ancienneté du directeur, nombre d'inspections, nombre de maîtres, nombre de toilettes, etc.). Il en est de même pour les incohérences liées au non-respect des filtres, qui ont été identifiées et corrigées pour tous les pays.

L'identification des valeurs atypiques/aberrantes a été contextualisée selon le pays. Ceci a parfois mené à reconsidérer la procédure statistique (méthode de la moyenne plus ou moins trois écarts-types) et à la fixer selon les réalités du pays.

Les rapports faisant état des incohérences identifiées ont été communiqués aux équipes nationales qui ont apporté les corrections nécessaires après consultation des instruments de collecte et parfois des directeurs des écoles concernées.

5.3. Mission de finalisation du nettoyage des données dans les pays

Bien que les rapports faisant état des incohérences identifiées dans les données aient été communiqués par le PASEC aux équipes nationales en vue de leur correction, une mission d'appui et de contrôle technique de cette opération a été effectuée par le PASEC dans la plupart des pays. Ces missions ont été déterminantes pour l'accélération du processus de nettoyage dans certains cas ou de sa finalisation dans d'autres.

5.4. Finalisation des bases de données : Recodage des items (non atteints et manquants par design)

À la fin du nettoyage des données, les bases de données des livrets A, B, C et D de la 5^e/6^e année de chaque pays ont été fusionnées en une seule base. Une recodification a été faite pour les items non atteints (code 8) des cahiers Élèves de fin de scolarité primaire et les données manquantes par design (code 7=non applicable). De même, les items non atteints des grilles de lecture des cahiers Élèves de 2^e année ont été recodés comme non applicables (code 7).

6. PONDÉRATION DES DONNÉES ET CALCUL DES POIDS ISSUS DE LA RÉPLICATION

Les données issues de l'évaluation ont été collectées à partir d'un échantillon et non auprès de la population complète d'écoles et d'élèves. Dès lors, il est impossible de connaître avec exactitude les paramètres de la population. On peut toutefois approcher la vraie valeur de ces paramètres à l'aide de méthodes statistiques appelées « pondération ». Le recours aux pondérations permet l'estimation des paramètres de population et de la variance d'échantillonnage afin d'étendre l'analyse de ces données à l'échelle nationale.

6.1 Pondération des données et calcul des taux de participation

6.1.1 Pondération des données

La pondération des données consiste à attribuer un « poids » à chaque école et à chaque élève de l'échantillon en fonction de sa probabilité de sélection dans l'échantillon et des éventuels ajustements. Ce « poids » permet de s'assurer que chaque école et chaque élève de l'échantillon représente un nombre approprié d'écoles et d'élèves dans la population totale d'écoles et d'élèves. Ainsi, la pondération vise à calculer des paramètres de population au départ d'un échantillon représentatif de cette même population.

Comme il a été indiqué au chapitre 3, l'échantillon de l'évaluation PASEC2019 est issu d'un tirage systématique proportionnel à la taille des écoles et d'un tirage aléatoire simple de classes et d'élèves. La probabilité finale de sélection d'un élève n'est pas égale d'une école à l'autre et entre l'échantillon initial et final pour les raisons suivantes :

- 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 une probabilité proportionnelle à la taille de l'école, une mesure de taille doit être incluse dans la base de sondage. Il peut arriver dans certains pays que le nombre d'élèves par niveau scolaire dans l'école ne soit pas complètement répertorié dans les statistiques nationales. 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. En outre, même si les statistiques nationales contiennent ces informations, des données sur les effectifs dans les écoles sont collectées par les administrateurs de tests et permettent de mettre à jour ces informations. 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 ;
- la non-réponse des écoles et des élèves : Certaines écoles peuvent refuser de participer à l'évaluation. De même, certains élèves au sein des écoles participantes peuvent refuser de participer ou être absents le jour de l'évaluation ;
- la variation de la taille des classes au sein des écoles : Les classes de même niveau au sein d'une école ne sont pas nécessairement de tailles strictement égales. Cette variabilité introduit aussi une variabilité des probabilités de sélection des élèves, et donc une variabilité supplémentaire des poids finaux des élèves.

La pondération des données est faite dans les bases de données des élèves pour tous les pays. Le poids final d'un élève comprend six composantes :

- le poids initial de l'école ;
- éventuellement, l'ajustement pondéral pour la non-réponse des écoles dans leurs strates respectives ;
- le poids de la classe au sein de l'école ;
- le poids initial de l'élève au sein de sa classe ;
- l'ajustement pour la non-réponse des élèves dans leurs classes respectives.

Le poids final d'un élève de 6^e année s'écrit donc de la manière suivante :

$$W_ELS_FIN_{ij} = W_SCH_INIT_j * ADJ_SCH_j * W_CLA_j * W_ELS_INIT_{ij} * ADJ_ELS_j$$

Où :

$W_ELS_FIN_{ij}$ correspond au poids final de l'élève i dans l'école j en 6^e année ;

$W_SCH_INIT_j$ correspond au poids initial de l'école j ;

ADJ_SCH_j correspond au facteur d'ajustement pour la non-réponse de l'école j ;

W_CLA_j correspond au poids de la classe au sein de l'école j . 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 ;

$W_ELS_INIT_{ij}$ correspond au poids initial de l'élève au sein de la classe ;

ADJ_ELS_j correspond au facteur d'ajustement pour la non-réponse des élèves.

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

La pondération des données de 2^e année suit globalement la même procédure que celle pour la 5^e/6^e année. Le poids initial de l'école est repris de la pondération des données de 5^e/6^e année mais en tenant compte du sous-échantillonnage des écoles. En effet, l'échantillon de la 2^e année est issu de l'échantillon de la 5^e/6^e année par un tirage aléatoire simple et en sélectionnant la moitié des écoles.

L'ajustement pour le sous-échantillonnage est appliqué au niveau des strates explicites. Le poids final d'un élève de début de scolarité primaire comprend donc six composantes :

- le poids initial de l'école pour les données de 5^e/6^e année ;
- l'ajustement pour le sous-échantillonnage ;
- l'ajustement pondéral pour la non-réponse des écoles ;
- le poids de la classe au sein de l'école ;
- le poids initial de l'élève au sein de sa classe ;
- l'ajustement pour la non-réponse des élèves.

6.1.2 Le poids initial de l'école

Le poids initial de l'école ($W_SCH_INIT_j$) est l'inverse de la probabilité de sélection de l'école dans l'échantillon. Dans le cadre du tirage systématique avec probabilité proportionnelle à la taille utilisé lors de l'évaluation PASEC2019, ce poids est calculé par strate et obtenu en appliquant la formule suivante :

$$W_SCH_INIT_j = \begin{cases} \left(\frac{N_h}{n_{e_s}} \right) * \left(\frac{1}{MOS_j} \right) & \text{si } MOS_j < \frac{N_h}{n_{e_s}} \\ 1 & \text{par ailleurs} \end{cases}$$

Où :

MOS_j correspond à la mesure de la taille de l'école j comme mentionné au chapitre 3 ;

N_h correspond au nombre total d'élèves dans la strate h . Il est calculé en faisant la somme des mesures de taille MOS_j des écoles dans la strate h ;

n_{e_s} correspond au nombre d'écoles à sélectionner dans la strate h .

Par exemple, pour un pays x avec deux strates dont le nombre total d'élèves est de 350 000 pour la première strate et 250 000 élèves pour la seconde strate. On souhaite échantillonner 180 écoles pour la première strate et 120 écoles pour la deuxième. le poids d'une école de la première strate dont la mesure de la taille est de 50 élèves est donc $(350\,000 / 180) * (1 / 50) = 38,89$. Un poids de 38,89 signifie qu'une école de la première strate représente une quarantaine d'écoles au niveau de la strate. Dans la même logique, le poids d'une école de 100 élèves dans la strate 2 est de $(250\,000 / 120) * (1 / 100) = 20,83$. Une école sélectionnée dans la seconde strate représente ainsi une vingtaine d'écoles au niveau de la strate.

6.1.3 L'ajustement pondéral pour la non-réponse des écoles

Au cours de la collecte des données, des écoles peuvent refuser de participer à l'enquête. Afin de tenir compte des écoles non participantes et non remplacées, des ajustements pour la non-réponse des écoles sont faits. 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 la contribution combinée des écoles participantes et des écoles qui auraient dû participer et la contribution des écoles participantes. Le calcul de l'ajustement pondéral pour la non-réponse des écoles est fait à l'intérieur des strates explicites définies lors de la conception du plan d'échantillonnage.

6.1.4 Le poids de la classe au sein de l'école

Lors de la collecte des données dans les écoles, une seule classe de 5^e/6^e année ou de 2^e année est tirée aléatoirement par les administrateurs de tests. Le nombre de classes ($NB_CLASSES_j$) de 5^e/6^e année ou de 2^e année de l'école j échantillonnée est indiqué dans la fiche de suivi de l'école. C'est ce nombre qui permet de calculer le poids de la classe au sein de l'école. Comme une seule classe est choisie, le poids de la classe au sein de l'école j se calcule comme suit :

$$W_CLA_j = \frac{1}{\frac{1}{NB_CLASSES_j}} = NB_CLASSES_j$$

Si l'école ne dispose que d'une classe pour un niveau donné, le poids de la classe au sein de l'école sera alors égal à 1.

6.1.5 Le poids initial de l'élève au sein de sa classe

Dans une école échantillonnée et au sein de la classe sélectionnée par l'administrateur de tests,

théoriquement 25 élèves en 5^e/6^e année et 16 élèves en 2^e année sont tirés parmi l'ensemble des élèves inscrits, qu'ils soient présents ou non le premier jour de l'administration des tests. Le nombre d'élèves échantillonnés (NB_SAMP_j) dans l'école j doit au plus être égal à 25 ou à 16 respectivement pour la 5^e/6^e année et la 2^e année.

L'ensemble des élèves de la classe (NB_ELS_j) au sein de l'école j sont également répertoriés dans la fiche de suivi de l'école, ce qui permet de calculer le poids initial de l'élève dans sa classe :

$$W_ELS_INIT_j = \frac{1}{\frac{NB_SAMP_j}{NB_ELS_j}} = \frac{NB_ELS_j}{NB_SAMP_j}$$

Par exemple, dans une école qui compte 65 élèves en 2^e année et 70 élèves en 5^e/6^e année et dans laquelle 16 élèves ont été échantillonnés en 2^e année et 25 en 5^e/6^e année, le poids d'un élève de 2^e année et de 5^e/6^e année est respectivement de $65 / 16 = 4,0625$ et $70 / 25 = 2,8$. Ce poids initial de l'élève dans sa classe permet de dire qu'un élève de 2^e année de cette école représente environ 5 de ses camarades et qu'un élève de 5^e/6^e année en représente environ 3.

6.1.6 L'ajustement pour la non-réponse des élèves

Le facteur d'ajustement pour la non-réponse des élèves permet de tenir compte des cas d'absence des élèves aux tests. Un élève absent est un élève échantillonné mais qui n'a participé à aucune session d'évaluation. Le nombre d'élèves absents dans une classe sélectionnée (NB_AB_j) au sein de l'école j est calculé à partir des informations consignées sur chaque élève dans la fiche de suivi des élèves.

Un élève est considéré comme présent lorsqu'il a passé au moins une section du test de compréhension de l'écrit ou une section du test de mathématiques. Le nombre d'élèves de l'école j présents (NB_PRE_j) est calculé en faisant la somme de tous les élèves présents aux tests.

Le facteur d'ajustement pour la non-réponse des élèves est calculé comme suit :

$$ADJ_ELS_j = \frac{NB_PRE_j + NB_AB_j}{NB_PRE_j}$$

6.1.7 Calcul des taux de participation des écoles et des élèves

Les taux de participation (avant ou après remplacement) sont calculés sans pondération puis avec pondération. Les procédures de calculs sont présentées dans les points suivants.

6.1.8 Taux de participation des écoles avant remplacement

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'écoles de l'échantillon initial qui ont ou auraient dû participer.

$$T_R_NP_EC_i = \frac{NB_S_P}{NB_S}$$

Où :

NB_S_P correspond au nombre d'écoles sélectionnées de l'échantillon initial qui ont participé sans prise en compte des écoles de remplacement ;

NB_S correspond au nombre total d'écoles sélectionnées qui ont ou auraient dû participer (180 en 5^e/6^e année et 90 en 2^e année dans le cas où il n'y a pas d'exclusion après échantillonnage).

Le taux de participation pondéré des écoles avant remplacement implique les mêmes enregistrements que précédemment mais prend en considération le poids initial de l'école et les effectifs (SIZE) de la classe tels qu'inscrits dans la base de sondage. Autrement dit, le taux de participation pondéré des écoles avant remplacement est égal à la somme des produits ($W_SCH_INIT * SIZE$) des écoles de l'échantillon initial qui n'ont pas été exclues et qui ont participé divisée 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.

$$T_R_P_EC_I = \frac{\sum_{k \in S} W_SCH_INIT_k \times SIZE_k}{\sum_{k \in (S \cup NP)} W_SCH_INIT_k \times SIZE_k}$$

Où :

S , correspond à l'ensemble des écoles de l'échantillon initial qui n'ont pas été exclues et qui ont participé ;

NP correspond à l'ensemble des écoles de l'échantillon initial qui n'ont pas été exclues et qui n'ont pas participé ;

$W_SCH_INIT_k$ correspond au poids initial de l'école k , qui est l'inverse de la probabilité de sélection de l'école ;

$SIZE_k$ correspond au nombre d'élèves de 5^e/6^e ou de 2^e année de l'école k .

6.1.9 Taux de participation des écoles après remplacement

Le calcul des taux de participation (non pondérés et pondérés) en tenant compte des écoles de remplacement suppose que ces dernières soient comptabilisées au niveau du numérateur au même titre que les écoles de l'échantillon initial non exclues et qui ont participé et qu'elles soient comptabilisées au niveau du dénominateur 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 école de remplacement n'interviennent plus au dénominateur.

Le calcul du taux de participation non pondéré avec les écoles de remplacement est le suivant :

$$T_R_NP_EC_{R1_R2} = \frac{NB_S_R_P}{NB_S_R}$$

Où :

$NB_S_R_P$ correspond au nombre d'écoles sélectionnées de l'échantillon initial plus les écoles de remplacement qui ont participé ;

NB_S_R correspond au nombre total d'écoles sélectionnées qui ont ou auraient dû participer en comptabilisant les écoles de remplacement et en supprimant les écoles remplacées.

Le calcul du taux de participation pondéré avec les écoles de remplacement s'effectue comme suit :

$$T_R_P_EC_{R1_R2} = \frac{\sum_{k \in SR} W_SCH_INIT_k \times SIZE_k}{\sum_{k \in (SR \cup NP)} W_SCH_INIT_k \times SIZE_k}$$

Où :

SR correspond à l'ensemble des écoles de l'échantillon initial et des écoles de remplacement qui n'ont pas été exclues et qui ont participé ;

NP correspond à l'ensemble des écoles de l'échantillon initial qui n'ont pas été exclues et qui n'ont pas participé ;

$W_SCH_INIT_k$ correspond au poids initial de l'école k ;

$SIZE_k$ correspond au nombre d'élèves de 5^e /6^e ou 2^e année de l'école k .

Tableau 6.1 Taux de participation non pondérés et pondérés des écoles

PAYS	Échantillon prévu		Échantillon réalisé		Taux de participation non pondéré sans écoles de remplacement (%)		Taux de participation pondéré sans écoles de remplacement (%)		Taux de participation non pondéré avec écoles de remplacement (%)		Taux de participation pondéré avec écoles de remplacement (%)	
	6 ^e	2 ^e	6 ^e	2 ^e	6 ^e	2 ^e	6 ^e	2 ^e	6 ^e	2 ^e	6 ^e	2 ^e
Bénin	232	116	230	116	99,1	100	99,1	100	99,1	100	99,1	100
Burkina Faso	400	150	377	131	92,5	84,9	93,8	87,5	94,3	87,3	95,6	90
Burundi	218	109	216	109	99,1	100	98,9	100	99,1	100	98,9	100
Cameroun	280	140	276	139	97,5	98,6	96,2	98,9	98,6	99,3	97,7	99,1
Congo	250	125	245	122	95,2	95,2	96,9	98,6	98	97,6	98,7	99,2
Côte d'Ivoire	180	90	179	87	99,4	96,7	99,5	97,1	99,4	96,7	99,5	97,1
Gabon	190	95	190	93	94,7	91,6	97,6	90,1	100	97,9	100	97,8
Guinée	180	90	179	87	98,3	93,3	97,6	94,1	99,4	96,7	99,4	96,8
Madagascar	265	135	265	134	97,7	97	97,2	97,8	100	99,3	100	99,6
Niger	286	150	284	143	96,5	92,7	97,6	92	99,3	95,3	99,3	95,7
RDC	267	90	246	79	79,1	70,1	81,7	72	92,1	87,8	92,1	90,2
Sénégal	180	90	180	87	100	95,6	100	95,9	100	96,7	100	97,1
Tchad	418	150	393	144	91,9	94,7	94,9	97,5	94	96	96,6	99,6
Togo	280	140	280	140	100	100	100	100	100	100	100	100

Taux de participation des élèves

Le taux non pondéré de participation des élèves se calcule comme suit :

$$T_R_NP_ELS = \frac{\sum_k NB_PRE_k}{\sum_k (NB_PRE_k + NB_AB_k)}$$

Où :

NB_PRE_k correspond au nombre d'élèves présents (participants à une des quatre sessions des tests) dans l'école k ;

NB_AB_k correspond au nombre d'élèves absents des tests dans l'école k .

Le taux de participation pondéré des élèves est calculé avec le poids initial de l'élève, qui tient compte du poids ajusté de l'école, du poids de la classe et du poids de l'élève dans la classe.

Tableau 6.2 : Taux de participation non pondéré et pondéré des élèves

PAYS	Nombre d'élèves échantillonnés		Nombre d'élèves participants		Nombre d'élèves exclus		Nombre d'élèves absents		Taux de participation non pondéré (%)		Taux de participation pondéré (%)	
	6 ^e	2 ^e	6 ^e	2 ^e	6 ^e	2 ^e	6 ^e	2 ^e	6 ^e	2 ^e	6 ^e	2 ^e
Bénin	3823	1654	3823	1654	0	0	0	0	100,0	100,0	100,0	100,0
Burkina Faso	6619	1932	6499	1884	12	9	108	39	98,4	98,0	98,6	97,9
Burundi	5039	1727	4908	1664	2	4	129	59	97,4	96,6	97,6	96,5
Cameroun	5151	2084	4723	1780	10	28	418	276	92,0	86,7	94,5	85,2
Congo	4062	1730	3925	1553	0	2	137	175	96,6	89,9	96,5	91,6
Côte d'Ivoire	3844	1352	3811	1332	1	0	32	20	99,2	98,5	99,1	98,5
Gabon	3006	1216	2930	1157	8	4	68	55	97,7	95,5	97,7	95,3
Guinée	2826	1086	2825	1086	0	0	1	0	100,0	100,0	100,0	100,0
Madagascar	4940	2113	4758	1883	6	7	176	223	96,4	89,4	97,1	91,2
Niger	5917	1845	5579	1730	0	0	338	115	94,3	93,1	94,8	93,3
RDC	4489	1062	4380	1050	0	0	109	12	97,6	98,9	97,8	98,7
Sénégal	3845	1345	3832	1341	0	1	13	3	99,7	99,8	99,7	99,70
Tchad	4932	1843	4824	1727	0	1	108	115	97,9	93,8	97,6	94,0
Togo	6206	2200	6117	2092	3	20	86	88	98,6	96,0	98,6	96,2

6.2 Calcul de la variance d'échantillonnage

L'estimation de statistiques de population à partir d'un échantillon d'individus comporte une certaine incertitude, ou risque d'erreur, liée à la variabilité qui existe entre les différents échantillons possibles. La variance d'échantillonnage est une quantification de l'incertitude due à l'échantillonnage.

Le plan d'échantillonnage doit être pris en considération pour l'estimation de la variance d'échantillonnage. En effet, 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 la variance d'échantillonnage, ce qui affecte les intervalles de confiance des paramètres de population estimés.

Les méthodes de réplification ou de ré-échantillonnage sont utilisées pour estimer les variances d'échantillonnage. Ces méthodes impliquent de nombreuses étapes de calculs en série. Ces méthodes de réplification reposent sur un même processus en trois étapes :

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

L'évaluation internationale PASEC2019 utilise la méthode de réplification de type Jackknife pour le calcul de la variance d'échantillonnage. Plus précisément, il s'agit de la variante de la méthode du Jackknife pour les échantillons stratifiés appelée Jackknife 2 (JK2).

Comme indiqué au chapitre 3, le tirage des échantillons par une procédure systématique avec une probabilité proportionnelle à la taille et l'utilisation de variables de stratification permettent de minimiser la variance d'échantillonnage.

La méthode Jackknife 2 pour un échantillon stratifié à plusieurs niveaux consiste à grouper par paires, au sein de chaque strate, les unités premièrement échantillonnées (les écoles) selon l'ordre dans lequel elles ont été sélectionnées. La méthode génère autant de réplifications dans la strate qu'il y a de paires d'écoles, ou « pseudo-strates ».

De façon technique, la méthode JK2 comprend les étapes suivantes :

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

6.2.1 La construction des zones de Jackknife, ou pseudo-strates

Pour construire les zones de Jackknife, les écoles participantes sont triées selon l'ordre dans lequel elles ont été sélectionnées dans l'échantillon. Dans cette liste ordonnée, une école de remplacement participante prend la place de l'école de la liste initiale qu'elle remplace.

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 de la 5^e/6^e année étant constitué de 180 écoles, 90 pseudo-strates sont construites, alors qu'il y en a 45 pour la 2^e année. Pour les pays où l'échantillon dépasse 180 écoles en 5^e/6^e année et 90 écoles en 2^e année, les pseudo-strates qui sont au-delà de la pseudo-strate 90 (ou 45 pour la 2^e année) sont fusionnées. Ainsi, la pseudo-strate 91 est fusionnée avec la pseudo-strate 1, la 92 avec la 2 et ainsi de suite.

6.2.2 La génération des poids de réplifications

Après la construction des pseudo-strates, il s'agit de générer les poids de réplifications. Les poids de réplifications sont simplement des poids calculés lors du processus de pondération des données. Ainsi, dans chaque pseudo-strate formée précédemment, un indicateur ui aléatoirement codé 0 ou 1 est attribué à chacune des deux écoles de telle sorte que l'une des écoles est codée 0 et l'autre, 1. Cet indicateur détermine si le poids des élèves échantillonnés dans cette école sera doublé (lorsque $ui=1$) ou mis à zéro (lorsque $ui=0$) pour créer la réplification de l'échantillon correspondant à la pseudo-strate.

Pour chaque pays, 90 réplifications sont produites en 5^e /6^e année et 45 en 2^e année. Dans le cas où, après la collecte et le nettoyage des données, un pays se retrouve avec moins de 180 écoles en 5^e /6^e année et moins de 90 écoles en 2^e année, 90 réplifications sont néanmoins construites pour la 5^e /6^e année et 45 pour la 2^e année.

Prenons par exemple un pays qui se retrouve avec 170 écoles après la collecte et le nettoyage des données : pour les 85 premières réplifications²⁸, le poids des élèves d'une des deux écoles de chaque pseudo-strate est mis à zéro et celui des élèves de l'autre école est doublé ; pour les 5 dernières réplifications, le poids des élèves de toutes les écoles participantes est inchangé.

Toutes ces étapes du JK2 sont appliquées directement sur la base de données des élèves, qui inclut les performances des élèves, les variables contextuelles et les pondérations des élèves. Ainsi, à chaque élève j de la base de données de 5^e /6^e année est attribué un vecteur de 90 poids de réplifications alors qu'en 2^e année c'est un vecteur de 45 poids de réplification qui est attribué.

Par exemple, pour l'élève j de la base de données de 5^e /6^e année, les poids de réplifications sont désignés par $rwgt_{1j}, rwgt_{2j}, rwgt_{3j}, \dots, wgt_{90j}$ et pour un élève k de la base de données de 2^e année, les poids de réplifications sont désignés par $rwgt_{1k}, rwgt_{2k}, rwgt_{3k}, \dots, wgt_{45k}$.

Ces poids de réplifications sont les transformations du poids final de l'élève j ($rwgt_{0j}$). Le poids de la h^e réplification est calculé comme suit :

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

Où :

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

Ainsi, k_{hj} vaut 2 ($u_i=1$) pour une des deux écoles de la pseudo-strate h et k_{hj} vaut 0 ($u_i=0$) pour la seconde école, c'est-à-dire que la pondération des élèves d'une des deux écoles est doublée, tandis qu'elle est ramenée à zéro pour l'autre école. Pour toutes les autres écoles n'appartenant pas à la pseudo-strate h , la pondération du poids des élèves est inchangée. La h^e réplification de l'échantillon contient donc autant d'élèves que l'échantillon principal.

²⁸ Pour les 170 écoles, on forme $170/2=85$ paires d'écoles ou pseudo-strates. Mais 90 réplifications doivent être formées. Ainsi, 5 réplifications seront ajoutées.

7. MISE À L'ÉCHELLE DES DONNÉES COGNITIVES

1. This document describes the outcomes of applying Item Response Theory (IRT) scaling and plausible value methodology to the PASEC assessment data.

7.1 Scaling model

2. The mixed coefficients multinomial logit model (MCMLM) as described by Adams, Wilson and Wang (1997) was used to scale the PASEC data, and implemented by ACER ConQuest® software (Adams et al. 2014).
3. The use of the MCMLM model which a generalised form of the Rasch model (Rasch, 1960) was proposed because:
 - Of all available item response theory models, it provides the strictest assessment of psychometric validity.
 - It supports the construction and validation of meaningful described proficiency scales.
 - The Rasch family as implemented in ACER ConQuest can be, and is, used to explore and control item position effects. Further, the model can be routinely applied in contexts that require multidimensional scaling.
 - It supports equating tests for the purposes of maintaining and monitoring the validity of trends. The latter was important for placing the results of 2019 assessments to the scale established in 2014 assessment

7.2 Item analysis and calibrations

4. The scaling of PASEC data with the MCMLM model was implemented for item calibration at national and international levels; as well as for the generation of student performance scores.
5. National item calibration was performed separately for each national country samples using weighted data. Initial international item calibration was conducted by scaling the pooled PASEC data set of equally weighted national samples (the sum of the weights was set to 1000 for each country). Item means were constrained to zero during the estimation process. The outcomes of the national and international item calibrations were used to make a decision about how to treat each item in each country. This means that an item may have been deleted from PASEC altogether if it has poor psychometric characteristics in more than five countries (a “dodgy” item); it may also have been discarded from the scaling in particular countries because there was a misprint or other flaw in this item.
6. When reviewing the item calibration results, particular attention was paid to the fit of the items to the scaling model, item discrimination and item-by-country interactions.

7.3 Item response model fit (weighted mean square MNSQ)

7. For each item parameter, the ACER ConQuest® fit mean square index (Wu, M. L., Adams, R. J., Wilson, M. R. & Haldane, S. (2007).) was used to provide an indication of the compatibility of the model and the data. For each student, the model describes the probability of obtaining the different item scores. It is therefore possible to compare the model prediction and what has been observed for one item across students. Accumulating comparisons across students gives an item-fit statistic. As the fit statistics compare an observed value with a predicted value, the fit is an analysis of residuals. In the case of the item weighted MNSQ fit statistics,

values near one are desirable. A weighted MNSQ greater than one is associated with a low discrimination index, meaning the data exhibit more variability than expected by the model, and an infit mean square less than one is associated with a high discrimination index, meaning the data exhibits less variability than expected by the model.

7.4 Discrimination coefficients

8. For each item, the correlation between the students' score and aggregate score on the set for the same domain and booklet as the item of interest was used as an index of discrimination. If p_{ij} (calculated as x_{ij}/m_i) is the proportion of score levels that student i achieved on item j , and $p_i = \sum_j p_{ij}$ (where the summation is of the items from the same booklet and domain as item j) is the sum of the proportions of the maximum score achieved by student i , then the discrimination is calculated as the product-moment correlation between p_{ij} and p_i for all students. For multiple-choice and short-answer items, this index will be the usual point-biserial index of discrimination.
9. The point-biserial index of discrimination for a particular category of an item is a comparison of the aggregate score between students selecting that category and all other students. If the category is the correct answer, the point-biserial index of discrimination should be higher than 0.20 (Ebel and Frisbie, 1986). They set out the following recommendations regarding the index of discrimination:

Magnitude	Comment	Recommended action for item
> 0.39	Excellent	Retain
0.30 – 0.39	Good	Possibilities for improvement
0.20 – 0.29	Mediocre	Need to check/review
0.00 – 0.20	Poor	Discard or review in depth
< -0.01	Worst	Definitely discard

Table 1: Point-Biserial Index Ranges

10. Non-key categories should have a negative point-biserial index of discrimination. The point-biserial index of discrimination for a partial credit item should be ordered, i.e. categories scored 0 should have a lower point-biserial correlation than the categories scored 1, and so on.

7.5 National reports

11. Item calibrations were performed separately for Grade 2, Grade 6 and Teachers data. In addition, for the Grade 6 and Teachers data booklet ID was added to the model to quantify booklet effect.
12. After the national scaling was completed, all available national item statistics were imported into the PASEC item database. International level item statistics were also included in this

database. This allowed summarising national level statistics and performing comparisons with the international and national item statistics for each of the Grade 2, Grade 6 and Teachers assessments.

13. For each country graphical summaries of descriptive statistics by item were generated and subsequently reviewed. For each item, descriptive statistics in tabular form and ICC (Item Characteristics Curve) graphs were produced. Additionally for each item the results of the DIF analysis and corresponding IES plots (Item Expected Score, which is the ICC for dichotomously scored items) by subgroups were generated.

Item 8									
item:8 (F8)									
Cases for this item 2247 Item-Rest Cor. 0.48 Item-Total Cor. 0.51									
Item Threshold(s): -0.74 Weighted MNSQ 0.98									
Item Delta(s): -0.74									
Label	Score	Count	% of tot	Pt Bis	t	sig	PV1Avg:1	PV1 SD:1	
1	0	172	7.65	-0.23	-11.1	0.000	-0.224	0.978	
2	0	154	6.85	-0.26	-12.82	0.000	-0.381	0.917	
3	0	191	8.5	-0.18	-8.52	0.000	0.020	1.065	
4	1	1677	74.63	0.48	26.03	0.000	1.113	1.107	
6	0	31	1.38	-0.22	-10.74	0.000	-1.762	1.345	
9	0	22	0.98	-0.08	-3.75	0.000	-0.218	0.726	

Table 2: Example of Item Statistics in Tabular Form

14. Table 2 **Erreur ! Source du renvoi introuvable.** shows as an example the item statistics for test item F8 which was part of the item report. The tabular item statistics include the following information:
 - **Item Delta** – item difficulty in logits
 - **Score** – shows score assigned to each response category (response 3 is correct response- scored 1)
 - **% of total** – percentage of the responses given in each category by all students that answered the question
 - **Weighted MNSQ** – item fit
 - **Pt Bis** - percentages for each category, category correlations with overall score (Pt Bis)
 - **Item-Rest Cor.** – equal to pt bis of a correct category response
 - **PV1Avg:1** - the average ability within each category
15. Two thousand two hundred and forty seven students have responded to item F8 in this country. The national delta (difficulty) is -0.74. 74.63% (facility) of the students have answered this item correctly. Item fit statistics is equal to 0.98 which indicate very good fit to the model.

16. The first column **Label** are the original response categories. This is a multiple-choice item and therefore, the responses are: 1=A, 2=B, 3=C, 4=D, 6='multiple response' and 9='missing'. The second column shows the score assigned to each response category. The correct response to this item is 3 (C).
17. The third and fourth columns in Table 2 list the number and percentage of students in each category. In this country, 1677 students (74.63%) gave the correct response.
18. The point-biserial correlations are presented in column five. This is the correlation between a response category coded as a dummy variable (a score of 1 for students that responded with the current code and a score of 0 for students in other response categories) and the total domain score. Correct responses should have positive correlations with the total score, incorrect responses negative correlations. In this case all of the incorrect responses have negative point-biserials and the key category has a positive point-biserial of 0.48.
19. The two last columns (**PV1Avg:1** and **PV1 SD:1**) show the average ability of students responding in each category and the associated standard deviation. The average ability is calculated by domain. If an item is functioning well the group of students that gave the correct response should have a higher mean ability than the groups of students that provided each of the incorrect responses. This is true for this item.
20. Within the context of Rasch modelling, an item was deemed to exhibit differential item functioning (DIF) if the response probabilities for that item could not be fully explained by the ability of the student and a fixed set of difficulty parameters for that item.
21. In DIF analysis the aim was *not* to find items that have a higher or a lower p-value for one group than another — since this may result from genuine differences in the proficiency levels of the two groups. The aim of the DIF analysis was to identify items that appear to be too difficult or too easy, after controlling for differences in the average proficiency across the two groups. In the PASEC Main Survey, gender and language of assessment (where applicable) were explored. The DIF analysis was performed using the multi-facet model of ACER ConQuest (Wu, Adams, Wilson and Haldane, 2007).
22. For the analysis of item performance an Item Characteristic Curve (ICC) was constructed and reported for each item.

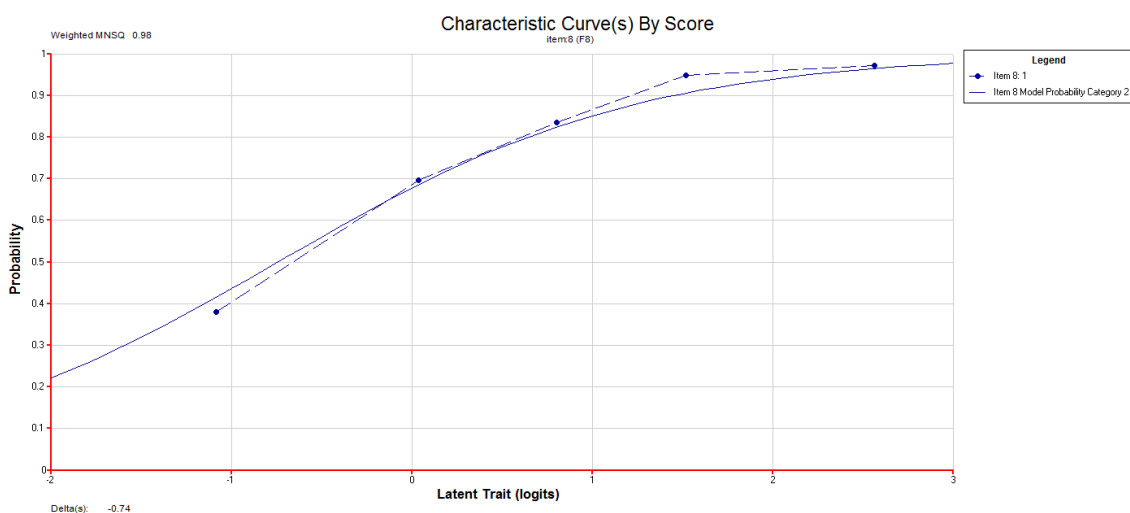


Figure 1: Example of Item Characteristic Curve

23. Figure 1 shows an example of an Item Characteristic Curve (ICC) for test item F8. With increasing ability (latent trait in logits) the probability of a correct response increases. The solid line shows the expected item responses given the IRT model and the dot points indicate the average percentage of correct responses in different ability groups. If the slope for the observed correct response is shallower than the expected curve, the item is less discriminating than predicted by the model; if it is steeper the item has a higher discrimination. In the example in Figure 1, the test item shows close to perfect fit to the scaling model.

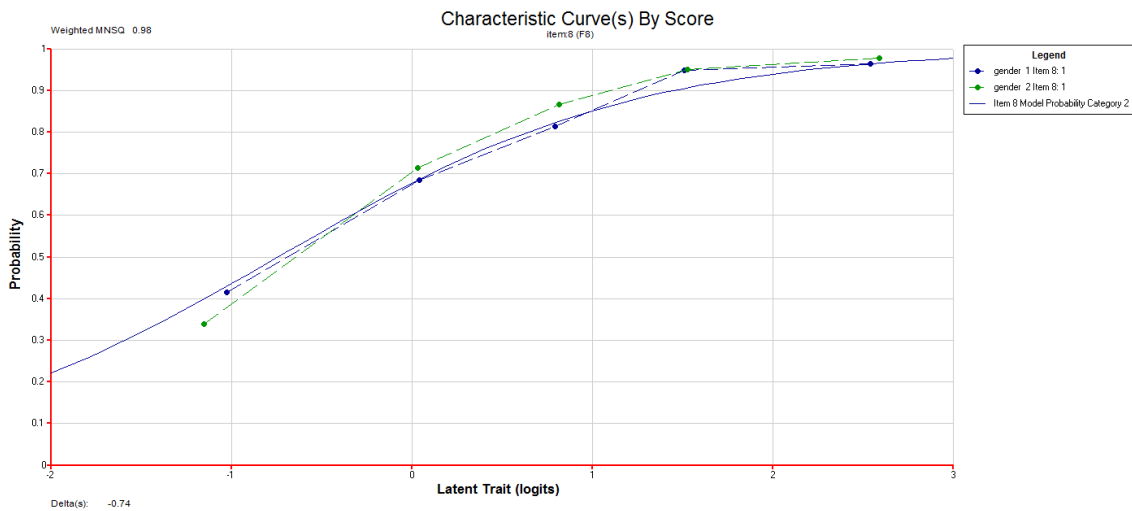


Figure 2: Item Expected Score Curve by Gender

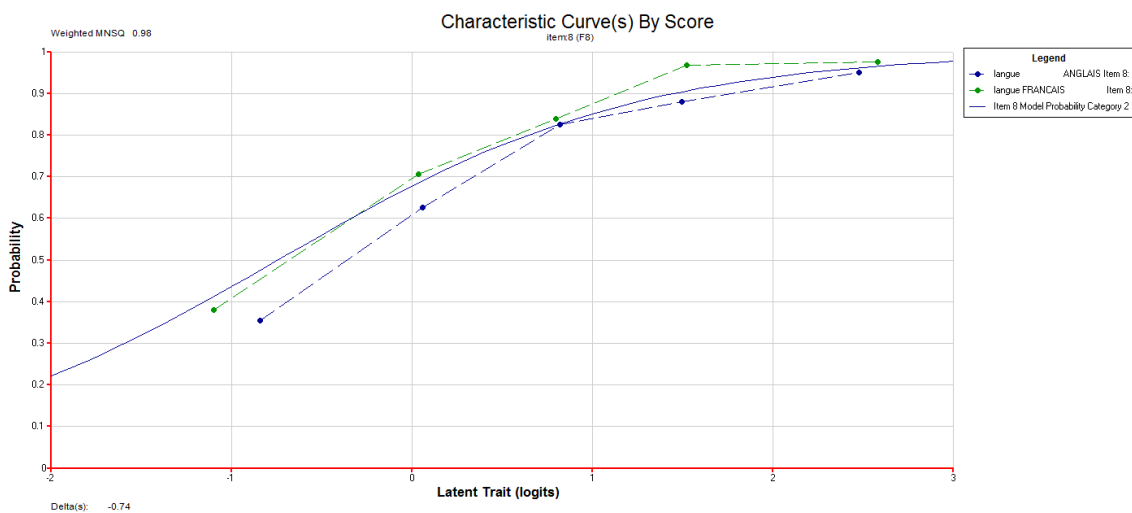


Figure 3: Item Expected Score Curve by Gender

24. Figure 2 and Figure 3 show an example of Item Expected Score (IES) by gender and by language for test item F8. The solid line shows the expected item responses given the IRT model and the green and blue dot points indicate the average percentage of correct responses in different ability groups for Boys and Girls in Figure 2, for English and French version of the item in Figure 3.
25. The IESs are a good graphical representation if the item was harder or easier for one of the groups of students. Figure 2 demonstrates that there was almost no difference in how difficult item F8 was for boys and girls. Figure 3 shows that the item was harder for students that were assessed in English than those assessed in French having the same abilities in the subject.

7.6 Item-by-country interaction

26. The national scaling provides nationally specific item parameter estimates. The consistency of item parameter estimates across countries was of particular interest. If the test measured the same latent trait per domain in all countries, then items should have the same relative difficulty or, more precisely, would fall within the interval defined by the standard error on the item parameter estimate (i.e. the confidence interval). Figure 4 illustrates item-by-country interaction plot for item F2. This item was harder than expected for Niger (NIG) and easier than expected for Burundi (BDI).

Item-by-country Interaction Plots

Blue line - International Item difficulty

Green Bars are confidence intervals of Item difficulty for each country

if green bar is below blue line then this Item is easier than expected in this country

if green bar is above blue line then this Item is harder than expected in this country

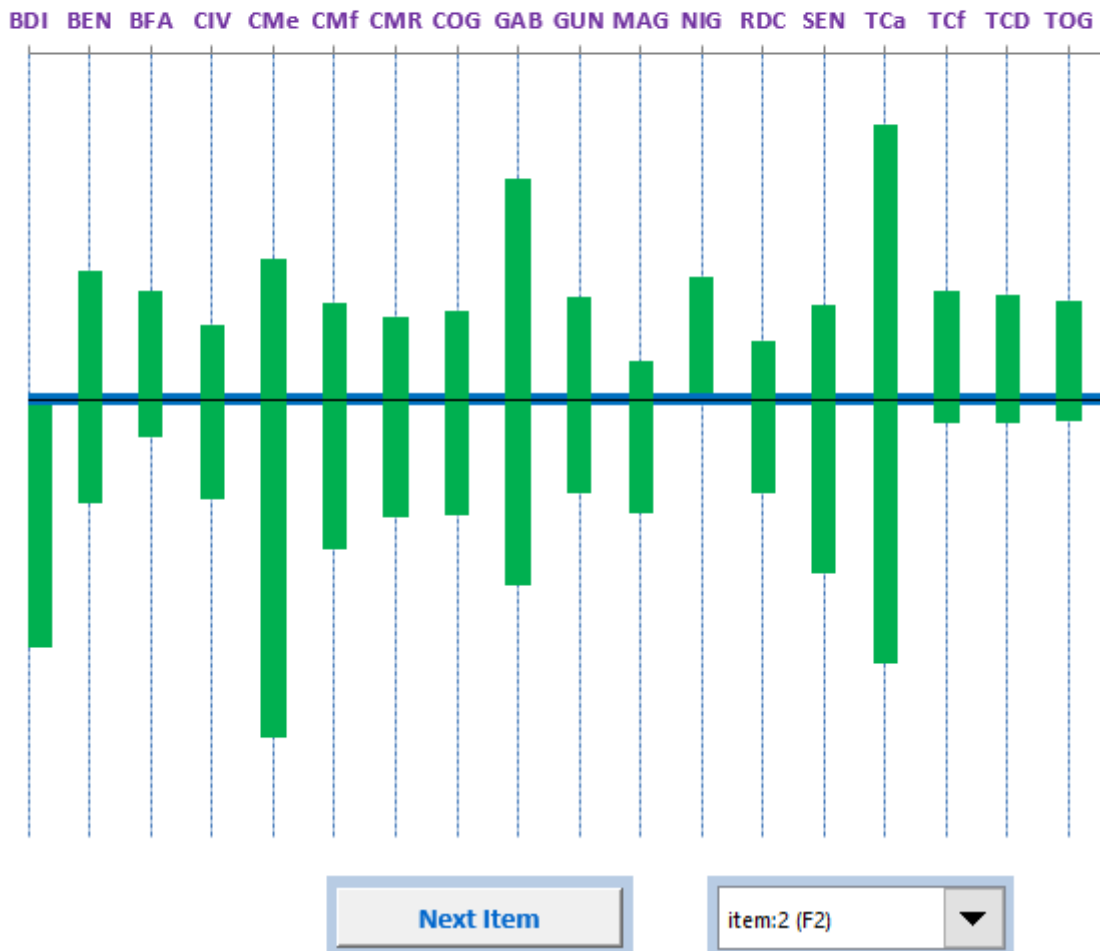


Figure 4: Item-by-country interaction

National List of Problematic Items Report

National List of Problematic Items lists for each country the “dodgy” items, i.e. the items that were flagged because:

- the item difficulty was significantly lower than expected;
- the item difficulty was significantly higher than expected;
- one of the non-key categories has a point bi-serial correlation higher than 0.05 (only reported if the category was chosen by at least 10 students);
- the key category has a negative point bi-serial;
- the item discrimination was lower than 0.2;
- the category abilities for partial credit items were not ordered (not applicable for Grade2);
- the fit is smaller or larger than expected.

27. Figure 5 contains a fragment of the National problematic items report that was generated for each country.

National List of Problematic Items

G6		R		Item by Country Interaction			Adjusted Correlation		Fit		Gender DIF
Benin FRE		No of	Easier	Harder	Non-key	Key	Low	Small	Large		
Item	Responses	Valid	than	than	PB is	PB is	adjusted	(high discr)	(low discr)		
			Expected	Expected	Positive	Negative	Correlation				
item:1 (F1)	1884										
item:2 (F2)	1884										
item:3 (F3)	1884										
item:4 (F4)	1883										
item:5 (F5)	1883										
item:6 (F6)	1883										
item:7 (F7)	1883										
item:8 (F8)	1883										
item:9 (F9)	1882			X							
item:10 (F10)	1882										
item:11 (F11)	1882										
item:12 (F12)	1882										
item:13 (F13)	1882			X							
item:14 (F14)	1882										
item:15 (F15)	1879										
item:16 (F16)	1879										
item:17 (F17)	1879										
item:18 (F18)	1879										
item:19 (F19)	1877										
item:20 (F20)	1876										
item:21 (F21)	1876										
item:22 (F22)	1876			X							
item:23 (F23)	1876										
item:24 (F24)	1859										
item:25 (F25)	1887										
item:26 (F26)	1887										
item:27 (F27)	1887										
item:28 (F28)	1887										
item:29 (F29)	1887						X		X		
item:30 (F30)	1887			X							
item:31 (F31)	1887										

Figure 5: Example of Item Statistics Shown in National Dodgy Item Report

28. The summary of the problematic items report shows for all countries items for which no indication of problems is evident, and items for which some such indication is present. A part of this report for Grade 6 assessment items is given in Figure 6.

International Summary of Problematic Items

Item	Item by Country Interaction		Adjusted Correlation		Fit		Gender DIF	
	Easier than	Harder than	Non-key PB is	Key PB is	Low adjusted	Small (high discr)	Large (low discr)	
Item	Expected	Expected	Positive	Negative	Correlation			
item:1 (F1)	0	2	1	0	0	0	0	1
item:2 (F2)	1	1	0	0	3	1	0	1
item:3 (F3)	0	1	3	0	1	0	0	0
item:4 (F4)	0	0	0	0	0	0	0	0
item:5 (F5)	0	2	2	0	2	0	0	0
item:6 (F6)	0	0	4	0	2	0	1	0
item:7 (F7)	0	1	1	0	2	0	0	0
item:8 (F8)	2	1	1	0	0	0	0	2
item:9 (F9)	0	1	1	0	1	0	3	0
item:10 (F10)	0	0	0	0	0	0	0	0
item:11 (F11)	0	1	4	0	2	0	0	0
item:12 (F12)	0	0	9	0	1	0	0	0
item:13 (F13)	0	1	0	0	1	0	0	0
item:14 (F14)	1	0	0	0	0	0	0	0
item:15 (F15)	1	0	1	0	0	0	0	0
item:16 (F16)	0	1	1	0	2	0	0	0
item:17 (F17)	1	0	0	0	0	0	0	0
item:18 (F18)	0	0	1	0	0	2	0	0
item:19 (F19)	0	1	1	0	1	0	0	1
item:20 (F20)	1	1	15	0	5	0	7	0
item:21 (F21)	0	0	1	0	2	0	0	0
item:22 (F22)	0	1	2	1	2	0	2	1
item:23 (F23)	0	1	1	0	1	0	2	0
item:24 (F24)	0	0	1	0	0	0	1	0
item:25 (F25)	0	1	0	0	0	0	0	0
item:26 (F26)	1	2	1	0	0	0	0	0
item:27 (F27)	0	0	0	0	0	0	0	1
item:28 (F28)	0	0	0	0	0	0	0	0
item:29 (F29)	0	2	0	0	2	0	5	0

Figure 6: Summary of problematic items

29. Those items flagged as having unsatisfactory psychometric characteristics were then reviewed in consultation with PASEC Technical Advisors. The consultations resulted in the deletion of a number of items at international levels in Grade 6 and Teachers assessments. Item **M72** was deleted internationally in Grade 6 dataset and items **MD7** and **MD10** were deleted in Teachers dataset. No items were deleted at international level for Grade 2 assessment.

The results of item calibration by language revealed inconsistencies in the translation of a number of items from French to English and Arabic. The students' (and teachers') performance on the identified items was compromised by mistakes in the translation and thus those items required to be removed from the data in the corresponding domains of knowledge.

30. Table 3 lists the items that were recoded to ‘not applicable’ for students and teachers who were assessed in English and Arabic.

Assessment	Domain	National Deletions
Grade 2	Maths	Cameroon (English): M61, M62, M63
	Language	--
Grade 6	Maths	Cameroon (English): M6, M20, M34 Chad (Arabic): M1, M2, M6, M12, M13, M15, M17, M18, M22, M23, M25, M37, M38, M40, M47, M48, M50, M51, M56, M57, M72, M74, M77, M78
	Language	Cameroon (English): F5, F20, F62, F73 Chad (Arabic): F6, F12, F26, F3,3 F34, F35, F36, F49, F63, F64, F81, F83, F86, F90, F93
Teachers	Maths	Chad (Arabic): M9, M13, M17, M47
	Didactic Maths	Chad (Arabic): MD2, MD3, MD5, MD6, MD8
	Language	--
	Didactic Language	--

Table 3: Items deleted after item calibration

7.7 International calibration

31. The final international calibration of items in Grade 2, Grade 6 and Teachers datasets was performed only using the data collected from participants assessed in French. Thus, in Grade 2 assessment data from Burundi, Madagascar, part of Niger and part of Cameroon where participants were assessed in languages different from French were not contributing to the international calibration. In Grade 6 and Teachers assessments data from part of Cameroon and part of Chad where participants were assessed in languages different from French were not contributing to the international calibrations.
32. Items were calibrated internationally separately in each domain, taking into account international deletions as described above.
33. The Grade 2 test was administered orally to the students by an interviewer in one set order. Four items included in the Grade 2 test were partial-credit items and there were no non-reached items. For Grade 2 ACER ConQuest® model statement for the calibration model was

$$\text{item} + \text{item} * \text{step}$$

34. This calibration model was used to estimate the international item parameters for Mathematics and Language for Grade 2 using the entire PASEC data set of equally weighted countries.
35. The Grade 6 test booklet design for the PASEC 2019 Main Survey is presented in Table 4. There are four clusters of Language and four clusters of Mathematics materials, rotated in four booklets as follows.

Booklet A	PL1	PL2	PM1	PM2
Booklet B	PL2	PL3	PM2	PM3
Booklet C	PL3	PL4	PM3	PM4
Booklet D	PL4	PL1	PM4	PM1

PL = Language / PM = Maths

Table 4: Grade 6 Test Booklet Design, PASEC 2019

36. The Teachers test booklet design for the PASEC 2019 Main Survey is presented in Table 5. There are four clusters of Language and four clusters of Mathematics materials and two clusters of Mathematics and Language didactic materials, rotated in four booklets as follows.

Booklet A	PL1	PL2	PDL	PM1	PM2	PDM
Booklet B	PL2	PL3	PDL	PM2	PM3	PDM
Booklet C	PL3	PL4	PDL	PM3	PM4	PDM
Booklet D	PL4	PL1	PDL	PM4	PM1	PDM

PL = Language / PM = Maths / PDL= Didactics Language / PDM= Didactics Maths

Table 5: Teachers Test Booklet Design, PASEC 2019

37. For the PASEC 2019 Grade 6 and Teachers tests designs it was expected that there would be booklet influences on the estimated proficiency distributions due to the different location of domains within each of the booklets.
38. Modelling the order effect in terms of item positions in a booklet or at least in terms of cluster positions in a booklet would result in a very complex model. For the sake of simplicity in the international scaling, the effect was modelled separately for each domain at the booklet level. To correct the student mathematics and language scores for the booklet effects one set of the internationally estimated booklet parameters was used. It is important to recognise that the sum of the booklet correction values is zero for each domain, so the application of the above correction does not change the country means or rankings.
39. When estimating the item parameters for Grade 6 and Teachers, booklet effects were included in the measurement model to prevent confounding item difficulties and booklet effects. For the ACER ConQuest® model statement, the calibration model was:

item + booklet.

40. The booklet parameter, formally defined in the same way as item parameters, reflects booklet difficulty. This calibration model was used to estimate the international item parameters for mathematics and literacy in Grade 6; and for mathematics, literacy, didactic mathematics and didactic language in Teachers.
41. Non-reached items were treated as non-administered at this stage. The booklet parameter, formally defined in the same way as item parameters, reflects booklet difficulty. This calibration model was used to estimate the international item parameters for each of the domains in Grade 6 and Teachers tests.
42. The booklet parameters obtained from this analysis were not used to correct for the booklet effect. Instead, a set of booklet parameters for the Grade 6 and for Teachers booklets was obtained by scaling the entire data set of equally weighted countries treating non-reached as incorrect answers and anchoring item parameters obtained in the model described above.
43. Table 6 contains estimated booklet effect in logits for Grade 6.

Grade 6	Domains	
Booklet	<i>Language</i>	<i>Mathematics</i>
A	-0.049	0.024
B	-0.031	-0.002
C	0.042	-0.033
D	0.038	0.012

Table 6: Grade 6 estimated booklet effects in logits

44. Table 7 contains estimated booklet effect in logits for Teachers.

Teachers	Domains			
Booklet	<i>Language</i>	<i>Mathematics</i>	<i>Didactic Language</i>	<i>Didactic Mathematics</i>
A	-0.040	-0.020	0.013	0.012
B	0.065	-0.039	0.010	-0.010
C	0.008	0.003	-0.008	0.009
D	-0.034	0.056	-0.015	-0.011

Table 7: Teachers estimated booklet effects in logits

45. Final item parameter estimates for each PASEC 2019 item are included in Appendices section. Appendix 1 and Appendix 2 contain item statistics for Grade 2 Language and Mathematics items correspondingly. Appendix 3 and Appendix 4 contain item statistics for Grade 6 Language and Mathematics items correspondingly. Appendix 5 to Appendix 8 contain item statistics for Teachers Language, Didactic Language, Mathematics and Didactic Mathematics items correspondingly.

7.8 Test targeting

46. Cognitive data for each of the domains were separately scaled to examine the targeting of the tests. We show the match between the international item difficulty distribution and the distribution of student achievement for the Grade 2 and Grade 6 assessments of language and mathematics, and Teachers assessment of language, mathematics, didactic language and didactic mathematics. The figures consist of two panels for Grade 2 and three panels for Grade 6 and for Teachers. The first panel (students) shows the distribution of students' Rasch-scaled achievement estimates. Students at the top end of this distribution have higher proficiency estimates than the students at the lower end of the distribution. The second panel (item difficulties) shows the distribution of Rasch-estimated item difficulties. For Grade 6 and Teachers the third panel show the distribution of booklet difficulties.
47. A test is well targeted if the average of item difficulties is about the same as the average of the students' abilities and the item difficulties are evenly spread across the ability distribution.
48. In Figure 7 to Figure 14, the distribution of student proficiency shown by Xs^i is matched to the distribution of item difficulty in the corresponding test. The figures are constructed so that when a student and an item are at the same location on the scale, then the student has a 50% chance of responding correctly to the item.

Terms in the Model (excl Step terms)			
+item			
5			
	X		
	X		
	X		
4	X		
	XX		
	XX		
	XXX		
3	XXX		
	XXX 29		
	XXXX 26		
	XXXXX		
2	XXXXXX 5 24		
	XXXXXX		
	XXXXXXXX 25		
	XXXXXXXX 20		
	XXXXXXXXXX 28 33		
1	XXXXXXXXXX 1 23 36		
	XXXXXXXX 11 19 22		
	XXXXXXXX 12		
	XXXXXXXX 4 10 15 27 34		
0	XXXXXXXX 35		
	XXXXXXXX 18 37		
	XXXXXX 3 17 30 31 39		
	XXXXXX 14 16 21		
	XXXXXX		
-1	XXXX 9		
	XXXXX 13		
	XXXX 7 8 32		
	XXXX		
-2	XXX		
	XX 2		
	X 38		
	X 40		
-3	X 6		
	X		
=====			
Each 'X' represents 105.0 cases			
=====			

Figure 7: Grade 2 Mathematics Item Difficulty and Achievement Distributions

49. Maths Grade 2 test is well targeted on the international level. Several more difficult items could be added to the test to improve various students' abilities coverage by the test.

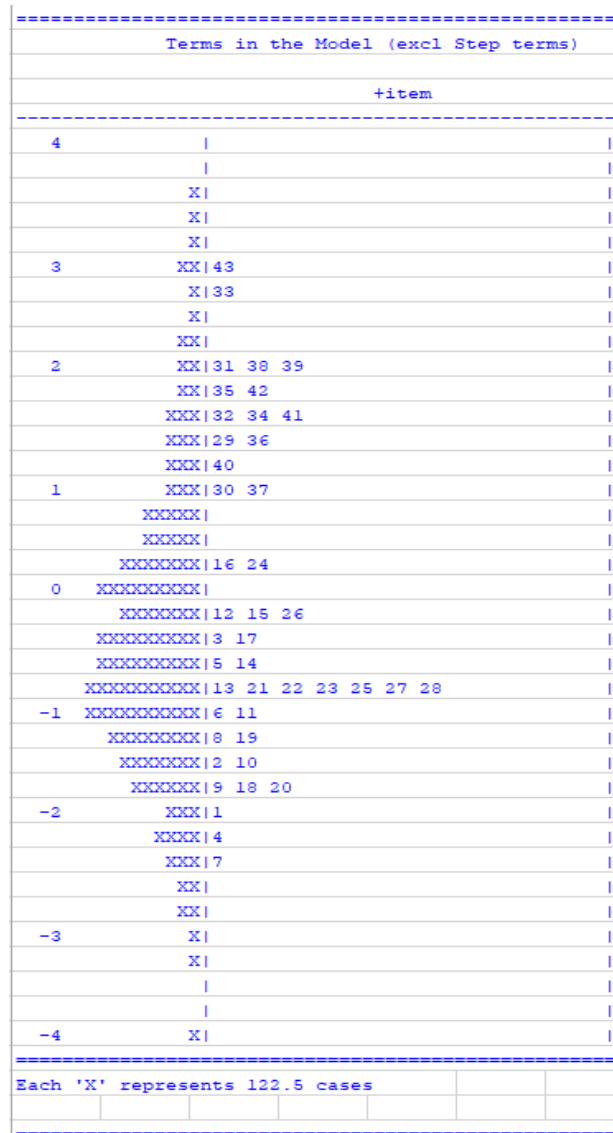


Figure 8: Grade 2 Language Item Difficulty and Achievement Distributions

50. For Grade 2 Language item distribution is not quite as good as for Grade 2 Maths, with some clustering of items. A number of items clustered in the area about -1 logits. These could be replaced with several more difficult and several less difficult items to improve coverage for various students' abilities. Overall, the Language test appear to be more difficult for the Grade 2 population than the Maths test.

Terms in the Model (excl Step terms)	
+item	+tid_livret
X	
X	
2 X	
XX 17 52	
XX 73	
X	
XX	
XX 4 45	
XXX 34	
XXX 12 55 67	
1 XXXX 53	
XXXXX 37 46 54	
XXXXX 28 76 78	
XXXXX 6 11 13 19 68 77	
XXXXXX 10 16 20 27 32 33 57 81	
XXXXXX 14 30 51	
XXXXXXXX 31 43 66 70 79	
0 XXXXXXXXXXX 3 15 25 42 83	1 2 4
XXXXXXXXX 38 40 56 59 65 71 74 75 82	3
XXXXXXXXX 26 47 50	
XXXXXXXX 7 49 62	
XXXXXXXX 21 41 61	
XXXXXXXXXX 2 39 44 84	
XXXXXXXXX 1 35	
XXXXXXXXX 29 63	
-1 XXXXXXXX 64	
XXXXX 18 22 24 48	
XXXXX 36 60 80	
XXXXX 8 9	
XXX 5 23	
XX	
XX 58	
X	
-2 X 69	
X	
Each 'X' represents 391.4 cases	

Figure 9: Grade 6 Mathematics Item Difficulty and Achievement Distributions

51. Test targeting is good for the Grade 6 Maths items.

Terms in the Model (excl Step terms)	
+item	+id_livret
X	
X	
X	
3 X	
XX	
XX	
XX	
XXX	
2 XXX	
XXXXX 20	
XXXX	
XXXXX 12	
XXXXX 36 42 63 83	
XXXXXX 5 22 29	
1 XXXXX 31 61 94 96	
XXXXXXXX 11 16 19 35 37 45 80 86 90	
XXXXXXXX 13 34 43 54 59 60 69 93	
XXXXXX 3 21 32 82	
XXXXXXXX 9 52 53 55 62 64 78 81 95	
XXXXXXXX 7 10 33 57 58 77 85 89	3 4
0 XXXXXXXX 6 30 44 51 67 68 87 91	1 2
XXXXXXXXXXXX 14 18 39 41 70 79 92	
XXXXXXXX 15 17 28 76 88	
XXXXXXXX 4 71	
XXXXXXXX 8 26 38 40 56 66 98	
-1 XXXXX 1 24 50 65 75 84	
XXXXXX 27 46 49 74	
XXXX 25 47 73	
XXX 48	
XX 23 72 97	
X	
-2 X 2	

Each 'X' represents 423.4 cases

Figure 10: Grade 6 Language Item Difficulty and Achievement Distributions

52. For Grade 6 Language there is a good symmetrical spread of items between student ability minus two to two, but no coverage at the top of the students' abilities range.

Terms in the Model (excl Step terms)						
	+item				+id_livret	
4						
	X					
	X					
	XX					
	XX					
	XX					
3	XX					
	XX					
	XXXX					
	XXXX					
	XXXXX					
	XXXXXXX					
2	XXXXX 37					
	XXXXXX 9 14 39 53					
	XXXXXX 19					
	XXXXXXX					
	XXXXXXXX 3 20 49 55					
	XXXXXXXXXX 59					
1	XXXXXXXX 33					
	XXXXXXXX 6					
	XXXXXXXX 25 46					
	XXXXXXXX 12 24 32					
	XXXXXXXX 1 22 30 40					
	XXXXXXXX 23 43					
0	XXXXXXXX 7 8 10 11 28				3 4	
	XXXXXXXX 18 38 50				1 2	
	XXXXX 44					
	XXXXX 2 13 31 35 36 41 45 54					
	XXXXXX 4 48 57 60					
	XXXXX 16 17 51 52 61					
	XXX 15 29 47					
-1	XXX 21 26					
	XX 5 34 42					
	X 56					
	X					
	X					
	X					
-2	27 62					
	58					
Each 'X' represents 108.2 cases						

Figure 11: Teachers Mathematics Item Difficulty and Achievement Distributions

53. For Teachers Mathematics test was easier than it should have been. The test items were clustered in the lower half of the teachers' mathematics abilities, thus having poor coverage of high range of teachers' abilities in Mathematics.

Terms in the Model (excl Step terms)	
titem	tid_livret
X	
4 X	
X	
X	
XX	
XX	
XXXX	
3 XXX	
XXXXX	
XXXXX	
XXXXX 24	
XXXXX	
XXXXXXXX 20	
2 XXXXXX 16 38	
XXXXXXXX	
XXXXXXXXXX 47	
XXXXXX 14	
XXXXXXXXXX 45	
XXXXXXXX 8 31 46	
1 XXXXXXXX 3 39	
XXXXXXXX 7 12 40	
XXXXXXXX 9	
XXXXXX 4 25 26 33 52	
XXXXXX 6 43	
XXXX 37	2
0 XXXXX 19 42 50	1 3 4
XXXX 11 22 29 30 36	
XXX	
XXX 1 17 18 27 44	
XX 2	
XX 32 34 41	
-1 X 23 51	
X 5 13 49	
X	
X	
15 28 48	
10 35	
-2	
21	

Each 'X' represents 122.9 cases

Figure 12: Teachers Language Item Difficulty and Achievement Distributions

54. For Teachers Language test was easier than it should have been. The test items were clustered in the lower half of the teachers' language abilities, thus having poor coverage of high range of teachers' abilities in Language. The test targeting in Language domain was worse than test targeting in Maths domain. Overall the Language test was too easy for the teachers.

Terms in the Model (excl Step terms)								
	+item							+id_livret
2								
	X							
	X							
	X							
	XX							
	X							
	XX							
	XX 2							
1	XXXX							
	XXXX							
	XXXXX							
	XXXXXX							
	XXXXX							
	XXXXXX 8							
	XXXXXX							
	XXXXXXX 6							
0	XXXXXXXX							
	XXXXXXXXXX 5							
	XXXXXXXX 1							
	XXXXXXXX 3 4							
	XXXXXXXX 7							
	XXXXXXXX							
	XXXXXXXX							
	XXXXXXXX							
-1	XXXXXXXX							
	XXXXXXXX							
	XXXX							
	XXXXX							
	XXXXX							
	XXX							
	XX							
	XX							
-2	XX							
	X							
	X							
	X							
	X							
Each 'X' represents 113.7 cases								

Figure 13: Teachers Didactic Mathematics Item Difficulty and Achievement Distributions

55. For Teachers Didactics Mathematics test there were not enough items to make a main domain. In practice for the construction of the reliable scale the number of items in the domain should include at least 20 or more items.

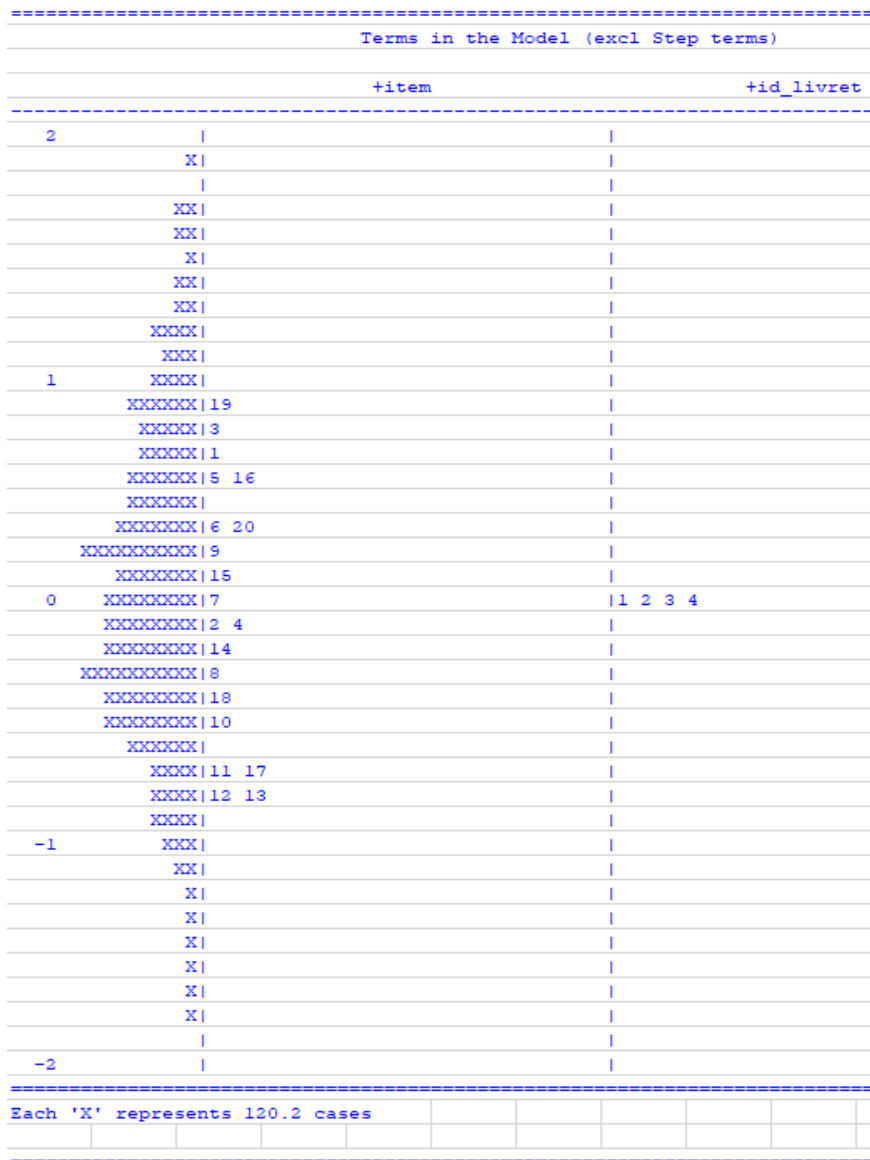


Figure 14: Teachers Didactic Language Item Difficulty and Achievement Distributions

56. For Teachers Didactics Language test the targeting of the test was good. We suggest developing several items in the lower and in the upper part of the scales.

7.9 Linking of PASEC 2019 to PASEC 2014 scale for Grade 2 and Grade 6

57. Reporting scales for Mathematics and Language domains in Grade 2 and Grade 6 were established in PASEC 2014 assessment. The design of PASEC 2019 included a number of items in both Mathematics and Language domains that were used in PASEC 2014 assessment. The sets of items that appeared in both 2014 and 2019 assessments are called common (or link) items. These sets of items were used for linking of PASEC 2019 Mathematics and Language to PASEC 2014 scales using standard common item equating methods.

58. In PASEC 2019 Grade 2 assessment there were 19 and 29 common items in Mathematics and Language parts correspondingly. In Grade 6 assessment, there were 40 and 43 common items in Mathematics and Language parts correspondingly.
59. The quality of common items to be used in the link was assessed by comparing the relative difficulties of the common items in each domain between the two assessments. To enable comparison of relative difficulties, the set of common items difficulties in each domain from each assessment cycle was centred at zero.
60. Differences between relative item difficulties were expected to be close to zero. A difference of more than half a logit was considered to be large and resulted in breaking the link for that item, meaning that the item would be regarded as two different items in 2014 and 2019 assessments.
61. After comparing relative difficulties of common items between two assessments, several items were removed from the link. All items were kept in Grade 2 Mathematics link, four items were removed from the link in Grade 2 Language (I031, I032, I081 and I111), two items removed from the link in Grade 6 Mathematics (M4 and M39) and one item removed from the link in Grade 6 Language (f62).
62. The difference between the means of the item parameter estimates for the remaining link items was calculated for PASEC 2014 and PASEC2019. This difference is the shift required to bring PASEC 2019 to the PASEC 2014 scale in each domain of Grade 2 and Grade 6. The estimated scale shifts for each domain are shown in
63. Table 8. These shift values were added to the estimates student performance scores in corresponding domains for all students in all countries.
64. The steps involved in linking the PASEC 2019 and PASEC 2014 language and mathematics scales for Grade 6 and Grade 2 were as follows:
- *Step 1:* International item parameter estimates for Language and Mathematics were obtained from the PASEC 2019 calibration.
 - *Step 2:* A shift constant was computed for each domain to place the above item parameters estimates on the PASEC 2014 scale so that the mean of the item parameter estimates for the common items was the same in 2019 as it was in 2014.
 - *Step 3:* The 2019 student performance scores (PVs) were estimated with item parameters anchored at their 2019 values.
 - *Step 4:* The above estimated student performance scores were transformed with the shift computed in step 2
65. In each case the transformation that equates the 2019 data with previous data depends upon the change in difficulty of each of the individual link item and as a consequence the sample of link items that have been chosen will influence the choice of transformation. This means that if an alternative set of link items had been chosen the resulting transformation would be slightly different. The consequence is an uncertainty in the transformation due to the sampling of the link items.

66. The uncertainty that results from the link-item sampling is referred to as linking error and this error must be taken into account when making certain comparisons between the results from different cycles of the assessment (Monseur and Berezner 2007). It is only possible to estimate the likely range of magnitudes for this error and take it into account when interpreting cross-cycle results. The likely range of magnitude for the linking errors is represented as a standard error. The values of the estimated link errors are summarised in Table 8 for each domain that was linked to the corresponding PASEC 2014 scale.

Assessment	Domain	Equating Shift	Link Error
Grade 2	READ (Language)	-0.115	0.049
	MATH (Mathematics)	0.130	0.033
Grade 6	READ (Language)	0.114	0.022
	MATH (Mathematics)	0.075	0.020

Table 8: Equating shifts for PASEC 2019

7.10 Scaling models and Reliability

67. The following three multi-dimensional scaling models have been used for producing students' and teachers' proficiency scores for the PASEC 2019 main study.

Grade 2

- Model 1: Language and Mathematics, used for reporting overall scores for *Language (READ)* and *Mathematics (MATH)* domains in Grade 2 students' assessment.

Grade6

- Model 2: Language and Mathematics, used for reporting overall scores for *Language (READ)* and *Mathematics (MATH)* domains in Grade 6 students' assessment.

Teachers

- Model 3: Language, Didactic Language, Mathematics and Didactic Mathematics, used for reporting overall scores for *Language (READ)*, *Didactic Language (RD)*, *Mathematics (MATH)* and *Didactic Mathematics (MD)* domains in Teachers' assessment.

68. The estimated reliabilities of the Main Survey tests for the main domains for the item pool over all participants in all countries for Grade 2, Grade 6 and Teachers assessments are given in Table 9. Non-reached items were treated as missing when calculating these reliability indices. The reliabilities reported in the table are from separate analysis of each domain. The Rasch-based reliability measures were provided by ACER ConQuest scaling software (Wu, Adams, Wilson & Haldane, 2007).

Assessment	Domain	Reliability
Grade 2	READ (Language)	0.87
	MATH (Mathematics)	0.87
Grade 6	READ (Language)	0.84
	MATH (Mathematics)	0.84
Teachers	READ (Language)	0.83
	MATH (Mathematics)	0.85
	RD (Didactic Language)	0.71
	MD (Didactic Mathematics)	0.56

Table 9: Main domains reliabilities by domain and by assessment

69. The reliability of Mathematic Didactic domain in Teachers assessment is very low – we do not recommend to report this scale.

7.11 Correlations between Main Domains

70. The latent correlations between the ability estimates for individual students in main domains, as estimated by ACER *ConQuest*[®] (Wu, Adams and Wilson, 1997), using international pooled data set with equally weighted countries are given in Table 10 for Grade 2, Grade 6 and Teachers assessments correspondingly. It is important to note that these latent correlations are unbiased estimates of the true correlation between the underlying latent variables. As such they are not attenuated by the un-reliability of the measures and will generally be higher than the typical product moment correlations that have not been de-attenuated for un-reliability.

Grade 2		LANG				
	MATH	0.82				
Grade 6		LANG				
	MATH	0.78				
Teachers		LANG	LD	MATH	MD	
	LANG					
	LD	0.70				
	MATH	0.72	0.59			
	MD	0.77	0.69	0.83		

Table 10: Correlations between main domains by assesment

71. The correlation between Mathematics and Didactic Mathematics domains in Teachers assessment is the highest of all cross domain correlations. The Didactic Mathematics scale

was not recommended to be reported on as a separate scale without significant improvement and increase of the number of items.

7.12 Student score generation

72. The imputation methodology usually referred to as plausible values (PVs) has been used to estimate likely proficiencies for students that attained each score. The likely proficiency levels were estimated for students from each participating country using multidimensional scaling methods. In those estimations the item parameters were anchored at their values estimated from the international item calibration. A set of student questionnaire responses along with some school information from the school principals' questionnaires were used as background information in the conditioning model for the generation of the plausible value sets. In particular, school location (urban/rural) and school type (public/private/communal), school stratification information (ID_STRATE) and school mean performance score along with participant's gender were used as direct regressors for Grade 2, Grade 6 and Teachers datasets to allow secondary analysis and comparison of results by categories of the variables mentioned. For Grade 6 and Teachers datasets, test form indicators were also used as direct regressors during scaling process.
73. All other responses from the corresponding student/teacher questionnaire instrument were represented in the scaling model by factors derived from a principal components analysis (PCA) to avoid co-linearity between the individual variables and reduce the number of regressors. The conditioning variables were prepared using procedures based on those used in the United States National Assessment of Educational Progress (Beaton, 1987) and in TIMSS (Macaskill, Adams and Wu, 1998).
74. The steps involved in the process of creation of scaling models:
 - *Step 1.* Each variable used in the scaling model was dummy coded. Some variables (e.g. booklet ID, gender, etc) were directly used as conditioning variables
 - *Step 2.* All the dummy coded variables except of those that were used as direct regressors were subjects of the PCA that produced a component scores for each student (a sufficient number of components to account for 99 per cent of the variance in the original variables).
 - *Step 3.* The item-response model was fit to each national data set and the national population parameters were estimated using item parameters anchored at their international location and conditioning variables derived from the national principal components analysis (Step 2) and direct regressors (Step 1).
 - *Step 4.* Five vectors of student performance scores (PVs) were drawn for each domain in the assessment using the corresponding multi-dimensional scaling model.
75. The student performance scores for each domain in Grade 6 and Teachers assessments were corrected for the booklet effect (not applicable to Grade 2) by adding constant values reported as in Table 6 and Table 7. In Grade 2 and Grade 6 student performance scores were equated to PASEC 2014 scale by adding corresponding constant shift values as reported in Table 8.

7.13 Constructing conditioning variables

76. All available student-level information, other than their responses to the items in the booklets, is used either as direct or indirect regressors in the conditioning model. The preparation of the variables for the conditioning proceeds as follows.
77. For Grade 6 and Techers dataset models, indicator variables for booklet ID were included as deviation contrast codes and were used as direct regressors. Each booklet was represented by one variable, except for reference Booklet A. Booklet A was chosen at random and was fixed for all counties. With the deviation contrast coding the sum of each column is zero, whereas for simple contrast coding, used frequently, the sum is one.
78. The contrast coding schemes used for Grade 2, Grade 6 and Teachers conditioning is described in Appendix 9, Appendix 10 and Appendix 11 correspondingly.
79. The deviation contrast codes resulted in an intercept in the conditioning model that is the grand mean of all students that responded to items in a domain if only the booklet is used as independent variable. This way, the imputation of abilities for students that did not respond to any mathematics or language items is based on information from all booklets that have items in a domain and not only from the reference booklet as in simple contrast coding.
80. Other direct variables in the regression are gender (and missing gender if there were any) and deviation contrast codes for strata with the largest strata as reference strata, and school performance indicator calculated from initial estimation of students weighted likelihood estimate (WLE) scores along with a variable representing a product of school type (public/private/communal) and school location (urban/rural) indicators from principals questionnaire. All other categorical variables from the student questionnaire were dummy-coded. These dummy variables were analysed in a principle component analysis. The number of component scores that were extracted and used in the scaling model as indirect regressors was country specific and explained 99% of the total variance in all the original variables. The direct and indirect conditioning variables described above were used as regressors in the multidimensional conditioned item-response models.
81. The multi-dimensional conditioned item-response models were fitted to each national data set and the national population parameters were estimated using item parameters anchored at their international location. The description of models which were fit to the data on a country level is given in earlier section.

Appendix 1: Grade 2 Language Item Parameters

Note: Item difficulty values were corrected to bring the scale to Grade 2 2014 Language scale

Item	N of Cases	Facility	Item-Rest Cor	Delta	Error	Step1	Step2	Step3	Step4
L011	16315	75.65	0.36	-1.933	0.017				
L012	16302	68.70	0.39	-1.568	0.017				
L013	16296	48.05	0.45	-0.440	0.016				
L014	16304	77.62	0.38	-2.159	0.018				
L021	16313	55.07	0.41	-0.777	0.016				
L022	16318	61.03	0.40	-1.106	0.016				
L023	16316	82.09	0.37	-2.525	0.019				
L024	16316	66.44	0.33	-1.422	0.017				
L031	16314	72.96	0.44	-1.850	0.017				
L032	16306	68.18	0.44	-1.545	0.017				
L033	16303	62.74	0.38	-1.208	0.016				
L034	16240	43.93	0.45	-0.165	0.016				
L035	16307	56.45	0.41	-0.832	0.016				
L041	16216	54.25	0.55	-0.689	0.016				
L042	16215	45.03	0.55	-0.209	0.016				
L043	16214	37.60	0.55	0.228	0.016				
L051	16271	50.77	0.55	-0.433	0.010	-0.854	-0.219	1.074	
L061	16287	74.40	0.39	-1.871	0.017				
L062	16293	66.52	0.46	-1.388	0.017				
L063	16286	73.70	0.43	-1.827	0.017				
L064	16284	60.49	0.43	-0.993	0.016				
L065	16274	60.41	0.45	-1.005	0.016				
L066	16282	61.31	0.44	-1.041	0.016				
L071	16250	40.91	0.67	0.081	0.009	-1.442	-0.059	0.190	1.312
L081	16306	59.83	0.45	-0.934	0.016				
L082	16302	46.65	0.54	-0.249	0.016				
L083	16303	60.57	0.50	-0.981	0.016				

L084	16302	58.28	0.54	-0.885	0.016
L091	16141	21.99	0.55	1.374	0.018
L092	16218	27.47	0.62	0.945	0.017
L093	16119	16.24	0.62	1.957	0.020
L094	16212	22.03	0.67	1.399	0.018
L095	16113	11.15	0.56	2.576	0.021
L096	16217	18.90	0.65	1.607	0.019
L097	16113	18.11	0.64	1.748	0.019
L098	16219	25.05	0.69	1.168	0.018
L101	16279	30.70	0.63	0.749	0.017
L102	16281	15.25	0.59	2.029	0.020
L103	16281	15.20	0.56	2.016	0.020
L111	16180	26.25	0.58	1.058	0.018
L112	16171	20.51	0.62	1.524	0.019
L113	16186	16.64	0.58	1.815	0.019
L114	16163	9.37	0.41	2.814	0.111

Appendix 2: Grade 2 Mathematics Item Parameters

Note: Item difficulty values were corrected to bring the scale to Grade 2 2014 Mathematics scale

Item	N of Cases	Facility	Item-Rest Cor	Delta	Error	Step1	Step2
M011	15960	42.66	0.64	1.122	0.012	-0.182	0.182
M021	16165	88.45	0.48	-2.194	0.022		
M022	16174	66.62	0.68	-0.314	0.017		
M023	16173	56.72	0.68	0.344	0.017		
M024	16151	25.03	0.55	2.324	0.018		
M031	16128	92.68	0.28	-2.783	0.024		
M041	16202	82.16	0.36	-1.455	0.020		
M042	16203	80.84	0.39	-1.390	0.019		
M051	16195	76.24	0.62	-0.977	0.019		
M052	16192	55.69	0.62	0.390	0.017		
M053	16191	47.01	0.56	0.922	0.017		
M054	16197	53.27	0.62	0.541	0.017		
M061	16205	79.93	0.54	-1.305	0.019		
M062	16205	70.28	0.56	-0.534	0.018		
M063	16197	53.74	0.60	0.501	0.017		
M071	16188	67.97	0.65	-0.416	0.018		
M072	16182	66.44	0.69	-0.303	0.017		
M073	16181	62.53	0.66	-0.047	0.017		
M074	16173	46.81	0.68	0.927	0.017		
M075	16173	35.95	0.62	1.550	0.017		
M081	16189	68.89	0.63	-0.432	0.018		
M082	16174	48.83	0.60	0.835	0.017		
M083	16165	46.40	0.65	1.001	0.017		
M084	16168	26.86	0.54	2.276	0.018		
M085	16163	34.93	0.60	1.680	0.017		
M086	16158	20.57	0.47	2.760	0.019		
M091	16181	56.42	0.63	0.375	0.017		

M092	16169	41.70	0.61	1.233	0.017
M093	16166	19.34	0.37	2.831	0.019
M101	16177	67.78	0.31	-0.343	0.017
M102	16178	64.95	0.38	-0.248	0.017
M103	16183	81.08	0.32	-1.408	0.020
M104	16180	39.73	0.30	1.368	0.017
M111	16191	54.34	0.42	0.476	0.017
M112	16195	58.63	0.38	0.212	0.017
M113	16194	45.55	0.38	1.017	0.017
M114	16192	64.43	0.46	-0.142	0.017
M121	16152	90.04	0.40	-2.406	0.022
M122	16145	63.82	0.32	-0.175	0.017
M123	16149	91.52	0.32	-2.614	0.111

Appendix 3: Grade 6 Language Item Parameters

Note: Item difficulty values were corrected to bring the scale to 2014 Grade 6 Language scale

Item	N of Cases	Facility	Item-Rest Cor	Delta	Error
F1	30866	72.72	0.42	-0.820	0.010
F2	30866	87.41	0.38	-1.975	0.012
F3	30866	49.06	0.42	0.494	0.010
F4	30866	66.34	0.53	-0.441	0.010
F5	30866	33.56	0.38	1.343	0.010
F6	30866	57.71	0.38	0.052	0.010
F7	30866	52.98	0.45	0.275	0.010
F8	30866	70.33	0.50	-0.670	0.010
F9	30866	52.95	0.31	0.334	0.010
F10	30866	54.53	0.57	0.220	0.010
F11	30866	40.41	0.35	0.979	0.010
F12	30866	29.23	0.39	1.642	0.010

F13	30866	44.23	0.42	0.743	0.010
F14	30866	61.31	0.57	-0.143	0.010
F15	30866	64.63	0.50	-0.337	0.010
F16	30866	41.23	0.38	0.877	0.010
F17	30866	62.83	0.45	-0.265	0.010
F18	30866	58.80	0.60	-0.079	0.010
F19	30866	40.59	0.49	0.882	0.010
F20	30866	23.35	0.25	2.061	0.011
F21	30866	47.95	0.43	0.520	0.010
F22	30866	35.79	0.42	1.215	0.010
F23	30865	81.71	0.37	-1.581	0.011
F24	30866	70.20	0.50	-0.843	0.011
F25	31354	78.00	0.48	-1.177	0.011
F26	31352	71.14	0.47	-0.749	0.010
F27	31354	77.25	0.50	-1.087	0.011
F28	31354	65.58	0.51	-0.361	0.010
F29	31354	37.25	0.32	1.196	0.010
F30	31354	56.44	0.47	0.115	0.010
F31	31354	37.84	0.30	1.108	0.010
F32	31354	48.76	0.55	0.542	0.010
F33	31354	56.21	0.37	0.205	0.010
F34	31354	46.47	0.38	0.670	0.010
F35	31354	40.81	0.31	0.961	0.010
F36	31354	33.00	0.29	1.491	0.010
F37	31354	41.93	0.51	0.904	0.010
F38	31354	70.20	0.53	-0.641	0.010
F39	31354	61.16	0.43	-0.095	0.010
F40	31354	71.58	0.60	-0.744	0.010
F41	31354	61.41	0.57	-0.164	0.010
F42	31354	34.21	0.41	1.388	0.010

F43	31354	44.91	0.58	0.700	0.010
F44	31354	56.69	0.56	0.121	0.010
F45	31354	42.53	0.46	0.888	0.010
F46	31354	76.33	0.49	-1.082	0.011
F47	31354	76.99	0.48	-1.163	0.011
F48	31014	80.52	0.45	-1.364	0.011
F49	31014	76.80	0.48	-1.108	0.011
F50	31014	72.02	0.53	-0.785	0.010
F51	31014	56.76	0.59	0.087	0.010
F52	31014	52.06	0.54	0.322	0.010
F53	31065	51.10	0.50	0.421	0.010
F54	31065	46.54	0.34	0.735	0.010
F55	31065	52.55	0.45	0.324	0.010
F56	31065	70.98	0.57	-0.722	0.010
F57	31065	55.97	0.53	0.142	0.010
F58	31065	54.14	0.47	0.272	0.010
F59	31065	43.65	0.43	0.809	0.010
F60	31065	44.27	0.48	0.672	0.010
F61	31065	37.42	0.23	1.143	0.010
F62	31065	48.64	0.36	0.477	0.010
F63	31065	31.30	0.33	1.508	0.010
F64	31065	51.65	0.31	0.418	0.010
F65	31065	71.60	0.52	-0.769	0.010
F66	31065	70.16	0.59	-0.660	0.010
F67	31065	57.41	0.51	0.091	0.010
F68	31065	58.52	0.57	-0.031	0.010
F69	31065	44.07	0.46	0.826	0.010
F70	31065	61.19	0.56	-0.157	0.010
F71	31065	69.15	0.44	-0.558	0.010
F72	31065	81.70	0.44	-1.556	0.011

F73	30597	75.29	0.26	-1.163	0.011
F74	30597	75.27	0.45	-1.006	0.011
F75	30597	72.58	0.56	-0.817	0.010
F76	30597	62.78	0.52	-0.261	0.010
F77	30597	52.44	0.47	0.299	0.010
F78	30597	51.94	0.33	0.375	0.010
F79	30597	59.87	0.54	-0.062	0.010
F80	30597	42.58	0.47	0.846	0.010
F81	30597	49.50	0.45	0.448	0.010
F82	30597	48.65	0.46	0.565	0.010
F83	30597	30.81	0.28	1.523	0.010
F84	30597	72.39	0.50	-0.795	0.010
F85	30597	53.14	0.42	0.260	0.010
F86	30597	39.49	0.35	0.982	0.010
F87	30597	58.85	0.44	-0.030	0.010
F88	30597	62.70	0.50	-0.239	0.010
F89	30597	53.86	0.57	0.222	0.010
F90	30597	39.69	0.38	0.995	0.010
F91	30597	56.97	0.60	0.040	0.010
F92	30597	59.32	0.58	-0.126	0.010
F93	30597	44.25	0.43	0.782	0.010
F94	30597	39.69	0.41	1.014	0.010
F95	30597	49.39	0.56	0.468	0.010
F96	30597	37.20	0.41	1.067	0.010
F97	30597	81.18	0.47	-1.535	0.011
F98	30597	68.59	0.45	-0.726	0.101

Appendix 4: Grade 6 Mathematics Item Parameters

Note: Item difficulty values were corrected to bring the scale to Grade 6 2014 Mathematics scale

Item	N of Cases	Facility	Item-Rest Cor	Delta	Error
M1	30962	60.87	0.36	-0.679	0.009
M2	30962	57.21	0.32	-0.452	0.009
M3	30962	43.12	0.19	0.170	0.009
M4	30962	21.70	0.38	1.414	0.010
M5	30962	76.05	0.42	-1.459	0.010
M6	30962	33.96	0.35	0.684	0.009
M7	30962	53.15	0.31	-0.272	0.009
M8	30962	72.06	0.34	-1.219	0.010
M9	30962	72.93	0.45	-1.272	0.010
M10	30962	37.71	0.32	0.550	0.009
M11	30962	34.46	0.36	0.670	0.009
M12	30962	25.32	0.15	1.175	0.010
M13	30962	35.31	0.25	0.667	0.009
M14	30962	38.85	0.43	0.436	0.009
M15	30962	44.38	0.33	0.148	0.009
M16	30962	37.47	0.35	0.518	0.009
M17	30962	16.14	0.32	1.880	0.010
M18	30962	68.37	0.40	-1.074	0.009
M19	30962	34.51	0.36	0.687	0.009
M20	30962	35.63	0.32	0.566	0.009
M21	30962	52.79	0.42	-0.343	0.009
M22	31393	68.14	0.42	-1.030	0.009
M23	31393	75.28	0.41	-1.451	0.010
M24	31393	67.84	0.37	-0.984	0.009
M25	31393	45.86	0.47	0.088	0.009
M26	31393	49.86	0.38	-0.119	0.009
M27	31393	38.18	0.30	0.464	0.009

M28	31393	32.24	0.48	0.760	0.009
M29	31393	62.67	0.46	-0.711	0.009
M30	31393	39.24	0.39	0.414	0.009
M31	31393	42.36	0.29	0.311	0.009
M32	31393	38.39	0.52	0.462	0.009
M33	31393	37.42	0.32	0.553	0.009
M34	31393	24.33	0.30	1.229	0.010
M35	31393	59.74	0.32	-0.581	0.009
M36	31393	71.93	0.28	-1.214	0.009
M37	31393	30.32	0.39	0.934	0.009
M38	31393	46.52	0.46	0.064	0.009
M39	31393	55.44	0.37	-0.459	0.009
M40	31393	49.33	0.43	-0.040	0.009
M41	31393	55.85	0.40	-0.401	0.009
M42	31393	45.59	0.28	0.109	0.009
M43	31106	45.24	0.27	0.214	0.009
M44	31106	57.35	0.27	-0.450	0.009
M45	31106	22.49	0.23	1.392	0.010
M46	31106	30.66	0.29	0.918	0.009
M47	31106	50.78	0.31	-0.101	0.009
M48	31106	69.31	0.47	-0.983	0.009
M49	31106	53.45	0.41	-0.215	0.009
M50	31106	51.20	0.26	-0.160	0.009
M51	31106	41.59	0.34	0.357	0.009
M52	31106	16.93	0.22	1.916	0.010
M53	31106	28.83	0.21	1.004	0.009
M54	31106	32.06	0.38	0.871	0.009
M55	31106	27.47	0.41	1.134	0.010
M56	31106	48.79	0.34	0.009	0.009
M57	31106	37.40	0.40	0.554	0.009

M58	31106	78.91	0.39	-1.640	0.010
M59	31106	47.59	0.32	0.040	0.009
M60	31106	70.74	0.45	-1.175	0.009
M61	31106	55.20	0.42	-0.334	0.009
M62	31106	52.67	0.22	-0.218	0.009
M63	31106	64.88	0.34	-0.827	0.009
M64	30704	66.75	0.34	-0.940	0.009
M65	30704	49.92	0.36	-0.050	0.009
M66	30704	42.28	0.39	0.290	0.009
M67	30704	27.79	0.33	1.196	0.010
M68	30704	37.69	0.43	0.594	0.009
M69	30704	83.30	0.32	-1.928	0.010
M70	30704	44.48	0.31	0.276	0.009
M71	30704	48.52	0.41	0.020	0.009
M73	30704	17.87	0.27	1.842	0.010
M74	30704	48.56	0.41	0.019	0.009
M75	30704	49.10	0.32	-0.019	0.009
M76	30704	33.88	0.41	0.738	0.009
M77	30704	35.21	0.48	0.692	0.009
M78	30704	33.76	0.40	0.776	0.009
M79	30704	44.55	0.54	0.197	0.009
M80	30704	70.59	0.41	-1.091	0.009
M81	30704	37.47	0.34	0.527	0.009
M82	30704	47.61	0.44	-0.013	0.009
M83	30704	44.17	0.32	0.143	0.009
M84	30704	58.47	0.41	-0.542	0.084

Appendix 5 Teachers Language Item Parameters

Item	N of Cases	Facility	Item-Rest Cor	Delta	Error
F01	9494	81.80	0.46	-0.544	0.020
F02	9494	83.03	0.48	-0.621	0.020
F03	9494	59.83	0.43	0.904	0.018
F04	9494	67.98	0.59	0.414	0.018
F05	9494	88.21	0.46	-1.165	0.021
F06	9494	69.70	0.53	0.273	0.019
F07	9494	62.69	0.52	0.721	0.018
F08	9494	54.72	0.48	1.146	0.018
F09	9494	64.88	0.41	0.591	0.018
F10	9494	93.05	0.37	-1.874	0.023
F11	9494	77.92	0.43	-0.267	0.019
F12	9494	59.97	0.32	0.874	0.018
F13	9494	86.92	0.39	-1.139	0.021
F14	9494	47.89	0.40	1.521	0.018
F15	9770	92.47	0.27	-1.734	0.022
F16	9770	41.92	0.27	1.900	0.018
F17	9770	81.93	0.41	-0.486	0.020
F18	9770	82.03	0.49	-0.564	0.020
F19	9770	74.80	0.43	0.008	0.019
F20	9770	34.84	0.41	2.152	0.018
F21	9770	95.79	0.29	-2.500	0.024
F22	9770	75.52	0.45	-0.142	0.019
F23	9770	86.89	0.41	-1.034	0.021
F24	9770	30.66	0.25	2.406	0.018
F25	9770	66.93	0.22	0.471	0.018
F26	9770	65.11	0.34	0.395	0.018
F27	9547	80.71	0.44	-0.482	0.020
F28	9547	91.56	0.40	-1.612	0.022

F29	9547	77.22	0.41	-0.226	0.019
F30	9547	75.60	0.55	-0.142	0.019
F31	9547	53.90	0.49	1.181	0.018
F32	9547	85.20	0.26	-0.889	0.021
F33	9547	67.27	0.38	0.457	0.018
F34	9547	86.09	0.46	-0.923	0.021
F35	9547	92.95	0.39	-1.867	0.023
F36	9547	77.96	0.32	-0.257	0.019
F37	9547	71.79	0.46	0.167	0.019
F38	9547	36.29	0.35	2.012	0.018
F39	9547	59.44	0.34	0.896	0.018
F40	9547	60.75	0.35	0.756	0.018
F41	9222	85.43	0.32	-0.894	0.021
F42	9221	74.46	0.48	-0.037	0.019
F43	9222	68.91	0.40	0.349	0.019
F44	9222	82.38	0.42	-0.576	0.020
F45	9222	52.09	0.39	1.286	0.018
F46	9222	57.03	0.52	1.064	0.018
F47	9222	47.65	0.21	1.544	0.018
F48	9222	91.48	0.38	-1.646	0.022
F49	9222	87.01	0.45	-1.145	0.021
F50	9222	74.48	0.44	-0.074	0.019
F51	9222	85.83	0.38	-1.083	0.021
F52	9222	67.66	0.52	0.433	0.139

Appendix 6 Teachers Didactic Language Item Parameters

Item	N of Cases	Facility	Item-Rest Cor	Delta	Error
FD1	19018	37.15	0.27	0.692	0.011
FD2	19018	52.81	0.39	-0.081	0.011
FD3	19018	36.38	0.31	0.708	0.011
FD4	19018	51.98	0.39	-0.033	0.011
FD5	19018	38.79	0.28	0.594	0.011
FD6	19018	43.43	0.31	0.365	0.011
FD7	19018	50.88	0.12	0.070	0.011
FD8	19018	56.43	0.26	-0.266	0.011
FD9	19018	47.88	0.12	0.193	0.011
FD10	19018	60.63	0.38	-0.465	0.011
FD11	19017	65.74	0.27	-0.662	0.011
FD12	19018	69.23	0.30	-0.844	0.012
FD13	19018	67.93	0.34	-0.770	0.011
FD14	19018	55.29	0.16	-0.196	0.011
FD15	19018	48.65	0.25	0.157	0.011
FD16	19018	40.44	0.30	0.514	0.011
FD17	19018	67.79	0.31	-0.757	0.011
FD18	19018	59.83	0.33	-0.402	0.011
FD19	19018	33.47	0.22	0.847	0.011
FD20	19018	43.64	0.27	0.335	0.049

Appendix 7 Teachers Mathematics Item Parameters

Item	N of Cases	Facility	Item-Rest Cor	Delta	Error
M1	9516	60.73	0.36	0.345	0.021
M2	9516	74.73	0.41	-0.447	0.022
M3	9516	41.74	0.35	1.374	0.021
M4	9516	77.90	0.26	-0.634	0.023
M5	9516	85.50	0.39	-1.283	0.025
M6	9516	50.06	0.47	0.924	0.021
M7	9516	65.83	0.44	0.061	0.021
M8	9516	64.71	0.47	0.147	0.021
M9	9516	31.50	0.45	1.876	0.021
M10	9516	65.73	0.54	0.072	0.021
M11	9516	65.09	0.38	0.104	0.021
M12	9516	55.80	0.31	0.601	0.021
M13	9516	70.55	0.34	-0.363	0.022
M14	9516	31.41	0.26	1.909	0.022
M15	9516	78.96	0.39	-0.860	0.024
M16	9807	79.44	0.35	-0.808	0.023
M17	9807	79.52	0.40	-0.751	0.023
M18	9807	69.20	0.32	-0.059	0.021
M19	9807	35.93	0.46	1.722	0.021
M20	9807	44.10	0.35	1.320	0.021
M21	9807	82.79	0.27	-1.002	0.024
M22	9807	59.74	0.57	0.442	0.021
M23	9807	64.13	0.53	0.259	0.021
M24	9807	58.41	0.49	0.531	0.021
M25	9807	55.31	0.48	0.693	0.021
M26	9807	81.38	0.40	-0.976	0.024
M27	9807	90.55	0.30	-1.955	0.029
M28	9807	64.29	0.51	0.136	0.021

M29	9807	80.77	0.47	-0.864	0.024
M30	9807	58.96	0.47	0.419	0.021
M31	9544	74.88	0.46	-0.400	0.022
M32	9544	57.77	0.32	0.521	0.021
M33	9544	46.31	0.43	1.093	0.021
M34	9544	85.80	0.34	-1.256	0.025
M35	9544	75.45	0.28	-0.413	0.022
M36	9544	74.85	0.51	-0.440	0.023
M37	9544	29.97	0.29	2.070	0.022
M38	9544	67.96	0.54	-0.038	0.022
M39	9544	33.28	0.48	1.859	0.021
M40	9544	59.02	0.42	0.476	0.021
M41	9544	73.20	0.48	-0.365	0.022
M42	9544	83.67	0.39	-1.173	0.025
M43	9544	62.99	0.55	0.248	0.021
M44	9544	70.13	0.44	-0.199	0.022
M45	9544	73.40	0.44	-0.422	0.023
M46	9544	52.56	0.47	0.779	0.021
M47	9248	80.95	0.26	-0.873	0.024
M48	9248	75.80	0.27	-0.614	0.023
M49	9248	43.49	0.28	1.311	0.021
M50	9248	69.24	0.41	-0.155	0.022
M51	9248	77.50	0.24	-0.674	0.023
M52	9248	78.43	0.43	-0.721	0.023
M53	9248	31.02	0.43	1.901	0.022
M54	9248	74.24	0.51	-0.405	0.023
M55	9248	40.68	0.38	1.380	0.021
M56	9248	85.71	0.43	-1.374	0.026
M57	9248	76.43	0.43	-0.614	0.023
M58	9248	92.74	0.27	-2.270	0.031

M59	9248	45.13	0.49	1.129	0.021
M60	9248	76.02	0.22	-0.587	0.023
M61	9248	77.36	0.50	-0.696	0.024
M62	9248	89.91	0.29	-2.011	0.176

Appendix 8 Teachers Didactic Mathematics Item Parameters

Item	N of Cases	Facility	Item-Rest Cor	Delta	Error
MD1	18927	49.49	0.21	-0.245	0.012
MD2	18927	22.52	0.19	1.148	0.013
MD3	18927	51.63	0.34	-0.383	0.012
MD4	18927	49.90	0.25	-0.345	0.012
MD5	18927	45.92	0.27	-0.146	0.012
MD6	18927	41.80	0.37	0.098	0.012
MD8	18927	54.54	0.31	-0.512	0.012
MD9	18927	35.78	0.31	0.384	0.031

Appendix 9 Grade 2 Contrast coding

ITEM NAME	VARIABLE CODING	CONTRAST CODING	REGRESSOR TYPE
Student Questionnaire			
qe22	Numeric value Missing	Two dummy variables: value – median 0 0 1	Factor
qe23	Responses: 1, 2 Missing	Two dummy variables with default value of '0' and - national mode = '-1' in both dummies - corresponding category= '1' (including missing)	Direct
qe24			Factor
qe25			Factor
qe26a			Factor
qe26b			Factor
qe26c			Factor
qe27	Responses: 1, 2, 3, 4 Missing	Four dummy variables with default value of '0' and - national mode = '-1' in both dummies - corresponding category= '1' (including missing)	Factor
qe28			Factor
qe29	Responses: 1, 2 Missing	Two dummy variables with default value of '0' and - national mode = '-1' in both dummies - corresponding category= '1' (including missing)	Factor
qe210			Factor
qe211			Factor
qe212			Factor
qe213			Factor
qe214			Factor
qe215			Factor
qe216			Factor
qe217			Factor
Principal Questionnaire			
qd17	Responses: 1, 2, 3 Missing	Variable qd31 was recoded into Urban/Rural indicator with values of 1 - (qd31=1,2) and 2 -(qd31=3,4). A variable SCHOOL was created as a permutation of all values in qd17 and	Direct

qd31	Responses: 1, 2, 3, 4 Missing	Urban/Rural indicator. The values of SCHOOL ranged from 1 to 6 and missing if any of the two initial variables had missing value. Up to six dummy variables were created based on the SCHOOL variable with default value of '0' and - national mode = '-1' in all dummies - corresponding category= '1' (including missing).	
qd18	Responses: 1, 2 Missing	Two dummy variables with default value of '0' and - national mode = '-1' in both dummies - corresponding category= '1' (including missing)	Factor
qd19			Factor
qd21			Factor
qd23			Factor
qd29			Factor
qd32a			Factor
qd32b			Factor
qd32c			Factor
qd32d			Factor
qd32e			Factor
qd32f			Factor
qd32g			Factor
qd32h			Factor
qd32i			Factor
qd32j			Factor
qd32k	Factor		
qd33	Factor		
Other variables			
ID_STRATE	Numeric value	The number of dummies equal to the number of strata in the national dataset minus 1 with default value of '0' and - national mode = '-1' in all dummies - corresponding category= '1' (including missing)	Direct

SCH_R_Index (ID_ECOLE)	WLE scores of student performance in maths estimated during national item calibration	Scale score computed as mean of WLE scores in maths of students having the same school ID (ID_ECOLE) excluding the student	Direct
SCH_M_Index (ID_ECOLE)	WLE scores of student performance in language estimated during national item calibration	Scale score computed as mean of WLE scores in language of students having the same school ID (ID_ECOLE) excluding the student	Direct

Appendix 10 Grade 6 Contrast coding

ITEM NAME	VARIABLE CODING	CONTRAST CODING	REGRESSOR TYPE
Student Questionnaire			
qe62	Numeric value Missing	Two dummy variables: value – median 0 0 1	Factor
qe63	Responses: 1, 2 Missing	Two dummy variables with default value of '0' and - national mode = '-1' in both dummies - corresponding category= '1' (including missing)	Direct
qe64a			Factor
qe64b			Factor
qe64c			Factor
qe64d			Factor
qe64e			Factor
qe64f			Factor
qe64g			Factor
qe65			Factor
qe66			Factor
qe67a			Factor
qe67b			Factor
qe67c			Factor
qe68			Factor
qe69a			Factor
qe69b			Factor
qe69c			Factor
qe69d			Factor
qe69e			Factor
qe69f			Factor
qe610			Factor
qe611a			Factor

qe611 b			Factor
qe612	Responses: 1, 2, 3, 4	Four dummy variables with default value of '0' and - national mode = '-1' in both dummies - corresponding category= '1' (including missing)	Factor
qe613	Missing		Factor
qe614	Responses: 1, 2 Missing	Two dummy variables with default value of '0' and - national mode = '-1' in both dummies - corresponding category= '1' (including missing)	Factor
qe615	Responses: 1, 2, 3 Missing	Three dummy variables with default value of '0' and - national mode = '-1' in both dummies - corresponding category= '1' (including missing)	Factor
qe616	Responses: 1, 2, 3, 4 Missing	Four dummy variables with default value of '0' and - national mode = '-1' in both dummies - corresponding category= '1' (including missing)	Factor
qe617			Factor
qe618			Factor
qe619			Factor
qe620			Factor
qe621a	Responses: 1, 2 Missing	Two dummy variables with default value of '0' and - national mode = '-1' in both dummies - corresponding category= '1' (including missing)	Factor
qe621 b			Factor
qe621c			Factor
qe621 d			Factor
qe621 e			Factor
qe621f			Factor
qe621g			Factor
qe622			Factor
qe623			Factor
qe624	Responses: 1, 2, 3, 4 Missing	Four dummy variables with default value of '0' and - national mode = '-1' in both dummies - corresponding category= '1' (including missing)	Factor

qe625a	Responses: 1, 2 Missing	Two dummy variables with default value of '0' and - national mode = '-1' in both dummies - corresponding category= '1' (including missing)	Factor
qe625b			Factor
qe625c			Factor
qe625d			Factor
qe625e			Factor
qe625f			Factor
qe625g			Factor
qe626			Factor
qe627			Factor
qe628			Factor
qe629			Factor
qe630			Factor
qe631	Factor		
qe632	Responses: 1, 2, 3, 4 Missing	Four dummy variables with default value of '0' and - national mode = '-1' in both dummies - corresponding category= '1' (including missing)	Factor
qe633	Responses: 1, 2 Missing	Two dummy variables with default value of '0' and - national mode = '-1' in both dummies - corresponding category= '1' (including missing)	Factor
qe634	Responses: 1, 2, 3 Missing	Three dummy variables with default value of '0' and - national mode = '-1' in both dummies - corresponding category= '1' (including missing)	Factor
qe635a	Responses: 1, 2 Missing	Two dummy variables with default value of '0' and - national mode = '-1' in both dummies - corresponding category= '1' (including missing)	Factor
qe635b			Factor
qe635c			Factor
qe635d			Factor

qe635 e		Factor
qe635f		Factor
qe635g		Factor
qe635 h		Factor
qe635i		Factor
qe635j		Factor
qe635k		Factor
qe636a		Factor
qe636 b		Factor
qe636c		Factor
qe636 d		Factor
qe636 e		Factor
qe636f		Factor
qe636g		Factor
qe636 h		Factor
qe636i		Factor
qe637a		Factor
qe637 b		Factor
qe637c		Factor
qe637 d		Factor
qe637 e		Factor
qe637f		Factor
qe637g		Factor
qe638		Factor

qe639	Responses: 1, 2, 3, 4, 5, 6 Missing	Six dummy variables with default value of '0' and - national mode = '-1' in both dummies - corresponding category= '1' (including missing)	Factor
qe640	Responses: 1, 2, 3, 4 Missing	Four dummy variables with default value of '0' and - national mode = '-1' in both dummies - corresponding category= '1' (including missing)	Factor
qe641			Factor
qe642			Factor
qe643			Factor
Principal Questionnaire			
qd17	Responses: 1, 2, 3 Missing	Variable qd31 was recoded into Urban/Rural indicator with values of 1 - (qd31=1,2) and 2 - (qd31=3,4). A variable SCHOOL was created as a permutation of all values in qd17 and Urban/Rural indicator. The values of SCHOOL ranged from 1 to 6 and missing if any of the two initial variables had missing value. Up to six dummy variables were created based on the SCHOOL variable with default value of '0' and - national mode = '-1' in all dummies - corresponding category= '1' (including missing).	Direct
qd31	Responses: 1, 2, 3, 4 Missing		
qd18	Responses: 1, 2 Missing	Two dummy variables with default value of '0' and - national mode = '-1' in both dummies - corresponding category= '1' (including missing)	Factor
qd19			Factor
qd21			Factor
qd23			Factor
qd29			Factor
qd32a			Factor
qd32b			Factor
qd32c			Factor
qd32d			Factor
qd32e			Factor
qd32f			Factor
qd32g			Factor
qd32h			Factor

qd32i			Factor
qd32j			Factor
qd32k			Factor
qd33			Factor
Other variables			
ID_STR ATE	Numeric value	The number of dummies equal to the number of strata in the national dataset minus 1 with default value of '0' and - national mode = '-1' in all dummies - corresponding category= '1' (including missing)	Direct
SCH_R _Index (ID_EC OLE)	WLE scores of student performance in maths estimated during national item calibration	Scale score computed as mean of WLE scores in maths of students having the same school ID (ID_ECOLE) excluding the student	Direct
SCH_M _index (ID_EC OLE)	WLE scores of student performance in language estimated during national item calibration	Scale score computed as mean of WLE scores in language of students having the same school ID (ID_ECOLE) excluding the student	Direct
ID_LIV RET	Responses: A, B, C, D	Three dummy variables with the following values - booklet A = '-1' '-1' '-1' - booklet B = '1' '0' '0' - booklet C = '0' '1' '0' - booklet D = '0' '0' '1'	Direct

Appendix 11 Teachers Contrast coding

ITEM NAME	VARIABLE CODING	CONTRAST CODING	REGRESSOR TYPE
Teacher Questionnaire			
qm1	Responses: 1, 2 Missing	Two dummy variables with default value of '0' and - national mode = '-1' in both dummies - corresponding category= '1' (including missing)	Direct
qm2	Numeric value Missing	Two dummy variables: value – median 0 0 1	Factor
qm3	Responses: 1,2,3,4,5,6,7,8,9,10,11,12 Missing	Twelve dummy variables with default value of '0' and - national mode = '-1' in both dummies - corresponding category= '1' (including missing)	Factor
qm4	Responses: 1,2,3,4,5,6,7,8 Missing	Eight dummy variables with default value of '0' and - national mode = '-1' in both dummies - corresponding category= '1' (including missing)	Factor
qm5	Responses: 1, 2, 3, 4, 5, 6 Missing	Six dummy variables with default value of '0' and - national mode = '-1' in both dummies - corresponding category= '1' (including missing)	Factor
qm6	Numeric value Missing	Two dummy variables: value – median 0 0 1	Factor
qm7	Responses: 1, 2 Missing	Two dummy variables with default value of '0' and - national mode = '-1' in both dummies - corresponding category= '1' (including missing)	Factor
qm8			Factor
qm9a			Factor
qm9b			Factor
qm9c			Factor
qm9d			Factor
qm9e			Factor

qm10	Numeric value Missing	Two dummy variables: value – median 0 0 1	Factor
qm11ci	Responses: 1, 2 Missing	Two dummy variables with default value of '0' and - national mode = '-1' in both dummies - corresponding category= '1' (including missing)	Factor
qm11cp			Factor
qm11ce1			Factor
qm11ce2			Factor
qm11cm 1			Factor
qm11cm 2			Factor
qm12	Numeric value Missing	Two dummy variables: value – median 0 0 1	Factor
qm13	Responses: 1, 2, 3, 4, 5, 6 Missing	Six dummy variables with default value of '0' and - national mode = '-1' in both dummies - corresponding category= '1' (including missing)	Factor
qm14	Numeric value Missing	Two dummy variables: value – median 0 0 1	Factor
qm15a	Responses: 1, 2 Missing	Two dummy variables with default value of '0' and - national mode = '-1' in both dummies - corresponding category= '1' (including missing)	Factor
qm15b			Factor
qm15c			Factor
qm15d			Factor
qm15e			Factor
qm15f			Factor
qm15g			Factor
qm15h			Factor
qm15i			Factor
qm15j			Factor
qm15k			Factor

qm15l			Factor
qm15m			Factor
qm15n			Factor
qm16			Factor
qm17	Responses: 1, 2, 3, 4 Missing	Four dummy variables with default value of '0' and - national mode = '-1' in both dummies - corresponding category= '1' (including missing)	Factor
qm18a	Responses: 1, 2 Missing	Two dummy variables with default value of '0' and - national mode = '-1' in both dummies - corresponding category= '1' (including missing)	Factor
qm18b			Factor
qm18c			Factor
qm18d			Factor
qm18e			Factor
qm19	Responses: 1, 2, 3 Missing	Three dummy variables with default value of '0' and - national mode = '-1' in both dummies - corresponding category= '1' (including missing)	Factor
qm20ci	Responses: 1, 2 Missing	Two dummy variables with default value of '0' and - national mode = '-1' in both dummies - corresponding category= '1' (including missing)	Factor
qm20cp			Factor
qm20ce1			Factor
qm20ce2			Factor
qm20cm 1			Factor
qm20cm 2			Factor
qm21	Numeric value Missing	Two dummy variables: value – median 0 0 1	Factor
qm22a			Factor
qm22b			Factor
qm23a			Factor

qm23b			Factor
qm24	Responses: 1, 2 Missing	Two dummy variables with default value of '0' and - national mode = '-1' in both dummies - corresponding category= '1' (including missing)	Factor
qm25a	Responses: 1, 2, 3,4, 5 Missing	Four dummy variables with default value of '0' and - national mode = '-1' in both dummies - corresponding category= '1' (including missing)	Factor
qm25b			Factor
qm25c			Factor
qm25d			Factor
qm25e			Factor
qm25f			Factor
qm26a	Responses: 1, 2 Missing	Two dummy variables with default value of '0' and - national mode = '-1' in both dummies - corresponding category= '1' (including missing)	Factor
qm26b			Factor
qm26c			Factor
qm26d			Factor
qm26e			Factor
qm26f			Factor
qm26g			Factor
qm26h			Factor
qm26i			Factor
qm26j			Factor
qm26k			Factor
qm26l			Factor
qm26m			Factor
qm26n			Factor
qm26o	Factor		
qm26p	Factor		
qm26q	Factor		

qm26r			Factor
qm27	Responses: 1, 2, 3 Missing	Three dummy variables with default value of '0' and - national mode = '-1' in both dummies - corresponding category= '1' (including missing)	Factor
qm28	Responses: 1, 2 Missing	Two dummy variables with default value of '0' and - national mode = '-1' in both dummies - corresponding category= '1' (including missing)	Factor
qm29	Numeric value Missing	Two dummy variables: value – median 0 0 1	Factor
qm30	Responses: 1, 2, 3, 4	Four dummy variables with default value of '0' and - national mode = '-1' in both dummies - corresponding category= '1' (including missing)	Factor
qm31	Missing		Factor
qm32	Responses: 1, 2 Missing	Two dummy variables with default value of '0' and - national mode = '-1' in both dummies - corresponding category= '1' (including missing)	Factor
qm33	Responses: 1, 2, 3 Missing	Three dummy variables with default value of '0' and - national mode = '-1' in both dummies - corresponding category= '1' (including missing)	Factor
qm34	Responses: 1, 2 Missing	Two dummy variables with default value of '0' and - national mode = '-1' in both dummies - corresponding category= '1' (including missing)	Factor
qm35	Responses: 1, 2, 3 Missing	Three dummy variables with default value of '0' and - national mode = '-1' in both dummies - corresponding category= '1' (including missing)	Factor

qm36	Numeric value Missing	Two dummy variables: value – median 0 0 1	Factor
qm37a	Responses: 1, 2 Missing	Two dummy variables with default value of '0' and - national mode = '-1' in both dummies - corresponding category= '1' (including missing)	Factor
qm37b			Factor
qm37c			Factor
qm37d			Factor
qm37e			Factor
qm37f			Factor
qm38	Responses: 1, 2, 3 Missing	Three dummy variables with default value of '0' and - national mode = '-1' in both dummies - corresponding category= '1' (including missing)	Factor
qm39a	Responses: 1, 2 Missing	Two dummy variables with default value of '0' and - national mode = '-1' in both dummies - corresponding category= '1' (including missing)	Factor
qm39b			Factor
qm39c			Factor
qm39d			Factor
qm39e			Factor
qm39f			Factor
qm40	Responses: 1, 2, 3, 4, 5, 6 Missing	Six dummy variables with default value of '0' and - national mode = '-1' in both dummies - corresponding category= '1' (including missing)	Factor
qm41	Responses: 1, 2 Missing	Two dummy variables with default value of '0' and - national mode = '-1' in both dummies - corresponding category= '1' (including missing)	Factor
qm42a			Factor
qm42b			Factor
qm42c			Factor
qm43			Factor
qm44a			Factor
qm44b			Factor

qm44c			Factor
qm44d			Factor
qm44e			Factor
qm44f			Factor
qm44g			Factor
qm45a			Factor
qm45b			Factor
qm46a	Responses: 1, 2, 3, 4 Missing	Four dummy variables with default value of '0' and - national mode = '-1' in both dummies - corresponding category= '1' (including missing)	Factor
qm46b			Factor
qm46c			Factor
qm46d			Factor
qm46e			Factor
qm46f			Factor
qm46g			Factor
qm46h			Factor
qm46i			Factor
qm46j			Factor
qm46k			Factor
qm46l			Factor
qm46m			Factor
qm46n			Factor
qm46o			Factor
qm46p	Factor		
Principal Questionnaire			
qd17	Responses: 1, 2, 3 Missing	Variable qd31 was recoded into Urban/Rural indicator with values of 1 - (qd31=1,2) and 2 - (qd31=3,4). A variable SCHOOL was created as a	Direct

qd31	Responses: 1, 2, 3, 4 Missing	permutation of all values in qd17 and Urban/Rural indicator. The values of SCHOOL ranged from 1 to 6 and missing if any of the two initial variables had missing value. Up to six dummy variables were created based on the SCHOOL variable with default value of '0' and - national mode = '-1' in all dummies - corresponding category= '1' (including missing).	
qd18	Responses: 1, 2 Missing	Two dummy variables with default value of '0' and - national mode = '-1' in both dummies - corresponding category= '1' (including missing)	Factor
qd19			Factor
qd21			Factor
qd23			Factor
qd29			Factor
qd32a			Factor
qd32b			Factor
qd32c			Factor
qd32d			Factor
qd32e			Factor
qd32f			Factor
qd32g			Factor
qd32h			Factor
qd32i			Factor
qd32j			Factor
qd32k	Factor		
qd33	Factor		
Other variables			
ID_STRAT E	Numeric value	The number of dummies equal to the number of strata in the national dataset minus 1 with default value of '0' and - national mode = '-1' in all dummies - corresponding category= '1' (including missing)	Direct

SCH_R_Index (ID_ECOLE)	WLE scores of student performance in maths estimated during national item calibration	Scale score computed as mean of WLE scores in maths of students having the same school ID (ID_ECOLE) excluding the student	Direct
SCH_M_Index (ID_ECOLE)	WLE scores of student performance in language estimated during national item calibration	Scale score computed as mean of WLE scores in language of students having the same school ID (ID_ECOLE) excluding the student	Direct
SCH_RD_Index (ID_ECOLE)	WLE scores of student performance in didactic maths estimated during national item calibration	Scale score computed as mean of WLE scores in didactic maths of students having the same school ID (ID_ECOLE) excluding the student	Direct
SCH_MD_Index (ID_ECOLE)	WLE scores of student performance in didactic language estimated during national item calibration	Scale score computed as mean of WLE scores in didactic language of students having the same school ID (ID_ECOLE) excluding the student	Direct
ID_LIVRET	Responses: A, B, C, D	Three dummy variables with the following values - booklet A = '-1' '-1' '-1' - booklet B = '1' '0' '0' - booklet C = '0' '1' '0' - booklet D = '0' '0' '1'	Direct

8. ÉLABORATION DES ÉCHELLES DE COMPÉTENCES DES ÉLÈVES

Des conseillers techniques du Programme d'Analyse des Systèmes Educatifs de la CONFEMEN (PASEC) et trois experts de l'université de Liège (les professeurs A. Fagnant, C. Monseur & P. Schillings) se sont réunis ces 4, 5 et 6 mai 2015 au Centre International d'Études Pédagogiques à Paris pour définir les niveaux de compétences pour les évaluations de deuxième et de sixième primaire, et ce, tant en lecture qu'en mathématiques.

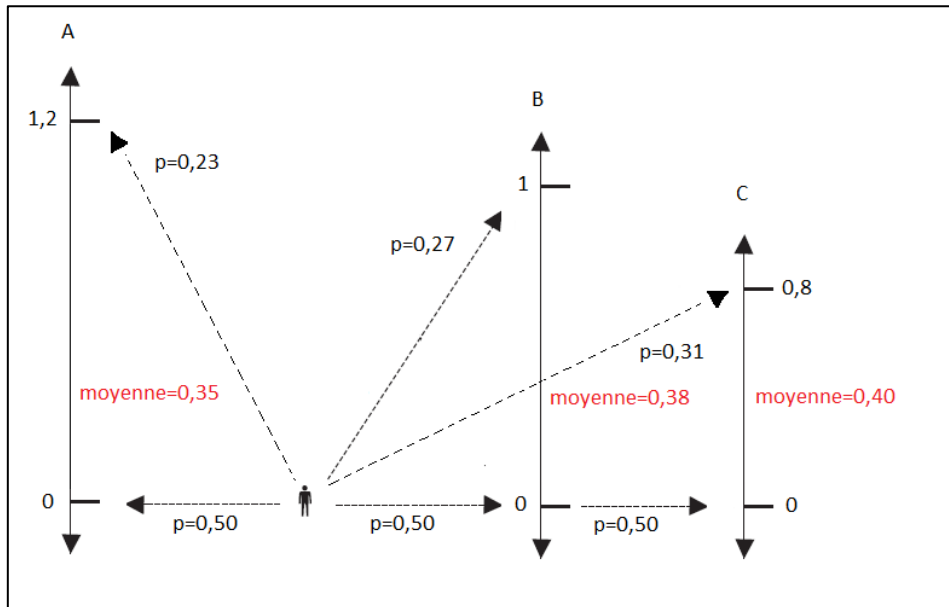
8.1. Méthodologie adoptée

Deux équipes d'experts ont été constituées. La première devait se charger des échelles de mathématiques et la seconde des échelles de compréhension de l'écrit. Ces équipes ont néanmoins interagi pour se coordonner au niveau de l'approche méthodologique et pour limiter, dans la mesure du possible, les différences dans les choix opérés.

Ainsi, les experts se sont entendus très rapidement sur la nécessité de définir, au sein d'une discipline et pour une année d'études donnée, des niveaux de compétences de longueur constante. En effet, comme l'illustre la figure 1, (i) la longueur d'un niveau de compétence, (ii) la *response probability*, qui correspond à la probabilité de réussite lorsque la compétence du sujet est identique à la difficulté de l'item et (iii) la probabilité moyenne de réussite d'un sujet donné à l'ensemble des items d'un niveau donné, sont interdépendants. Imaginons un individu de compétence 0.00 qui se situe au plus bas d'un niveau de compétence. Dans la figure 1, trois longueurs de niveau de compétence sont illustrées, à savoir (i) 1,20, (ii) 1,00 et enfin (iii) 0,80. Les probabilités que cet individu réussisse un item situé sur le point de césure supérieur du niveau de compétences sont respectivement de (i) 0,23, (ii) 0,27 et enfin (iii) 0,31. Il s'en suit que les probabilités moyennes (que l'on peut aussi estimer par la probabilité que l'individu a par rapport à un item situé au milieu du niveau de compétence) sont respectivement de (i) 0,35, (ii) 0,38 et enfin 0,40.

La figure 1 illustre parfaitement l'interdépendance entre longueur d'un niveau de compétence, et probabilité moyenne de réussite. Varier les longueurs des niveaux au sein d'une seule et même échelle ne permet pas de décrire de manière univoque ce qu'un élève d'un niveau donné est ou n'est pas capable de faire.

Figure 1 : probabilité de réussite et longueur d'un niveau de performance



Par ailleurs, les équipes se sont également entendues sur l'intérêt d'essayer de maintenir une longueur constante pour les deux disciplines et pour les deux années d'études. Cependant, ce dernier souhait n'a pas pu être rencontré en raison du faible nombre d'items en deuxième année comparativement à la sixième année.

Le premier travail des experts a consisté à classer les items par ordre croissant d'indices difficulté obtenus lors de la calibration internationale des données cognitives. Pour rappel, ces paramètres d'items résultent d'une modélisation IRT unidimensionnelle logistique à un paramètre en accordant un poids identique aux dix pays participants à cette première évaluation groupée du PASEC. Par défaut, cette calibration internationale a choisi comme *Response Probability* 0,50.

Les experts ont ensuite examiné sur la base du cadre de référence les processus cognitifs évalués, les contenus et le contexte des tâches (type de texte, longueur de texte...) et ont recherché des groupes de questions qui présentent une certaine homogénéité conceptuelle quant à la variable latente mesurée.

Ensuite, ils ont déterminé la longueur du niveau de compétence et le premier point de césure (les autres étant automatiquement définis par la contrainte adoptée de choisir des longueurs constantes) qui altèrent le moins l'homogénéité conceptuelle des tâches requises pas les items d'un niveau donné.

La dernière étape consiste à définir ce que signifie pour un élève d'être situé à un niveau donné. Deux alternatives ont largement été discutées et leurs avantages et inconvénients respectifs énumérés.

La première solution s'apparente à la définition adoptée par le Programme International de Suivi de l'Elève (PISA) de l'OCDE. En effet, « *The approach adopted for PISA 2000 was that it would only be useful to regard students as having attained a particular level if this would mean that we can have certain expectations about what these students are capable of in general when they are said to be at that level. It was decided that this expectation would have to mean at a minimum that students at a particular level would be more likely to solve tasks at that level than to fail them. By implication, it must be expected that they would get at least half of the items correct on a test composed of items uniformly*

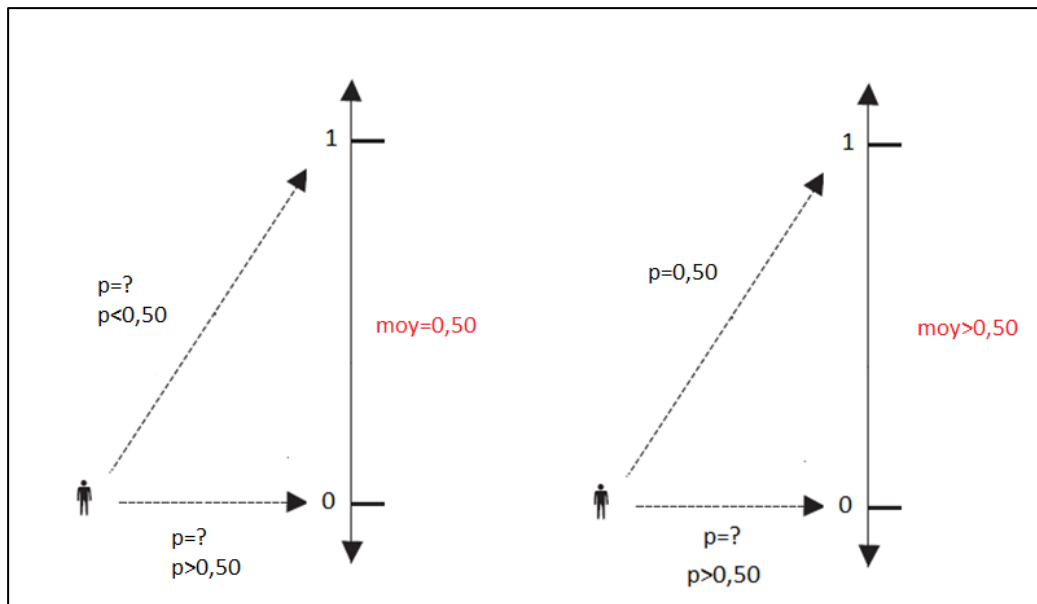
*spread across that level, which is useful in helping to interpret the proficiency of students at different points across the proficiency range defined at each level*²⁹. Un étudiant qui se situe au bas d'un niveau de compétence doit donc réussir au moins la moitié des items qui relèvent de ce niveau. Autrement dit, cet étudiant doit avoir une probabilité moyenne de réussite égale ou légèrement supérieure à 0,50.

Il importe donc de modifier la *Response Probability* pour qu'un individu situé sur le point de césure inférieur du niveau de compétence ait une probabilité moyenne 0,50 de réussir des items uniformément distribués sur ce même niveau de compétence. Cette situation est représentée graphiquement dans la partie gauche de la figure 2. Comme on peut le constater, puisque la probabilité moyenne est de 0,50, un élève situé sur le point de césure inférieur présente, pour environ la moitié des items, une probabilité de réussite inférieure à 0,50.

La seconde approche consiste à définir une *Response Probability* de sorte qu'un élève situé sur le point de césure inférieur d'un niveau ait une probabilité de 0.50 de réussir un item qui se situerait sur le point de césure supérieur du niveau. Dans ce cas, la probabilité moyenne sera supérieure à 0.50

Si la seconde approche est pédagogiquement plus défendable, elle conduit automatiquement à réduire les pourcentages d'élèves dans les niveaux supérieurs et à augmenter les pourcentages d'élèves dans les niveaux inférieurs. Pour ces raisons, il a été décidé d'opter pour la première alternative, à savoir un élève situé en bas d'un niveau de compétence doit avoir une probabilité moyenne de 0,50 de réussir les items distribués uniformément sur tout le niveau de compétence.

Figure 2 : deux approches pour définir ce que signifie pour un élève d'être situé à un niveau de compétence



L'étape suivante consiste à calculer la *Response Probability*. Quelle devrait être la probabilité de réussite d'un individu de compétence 0.00 de réussir un item de difficulté 0.00 pour que sa probabilité moyenne aux items uniformément distribués sur le niveau de compétence de longueur 1 soit égale à 0,50 ?

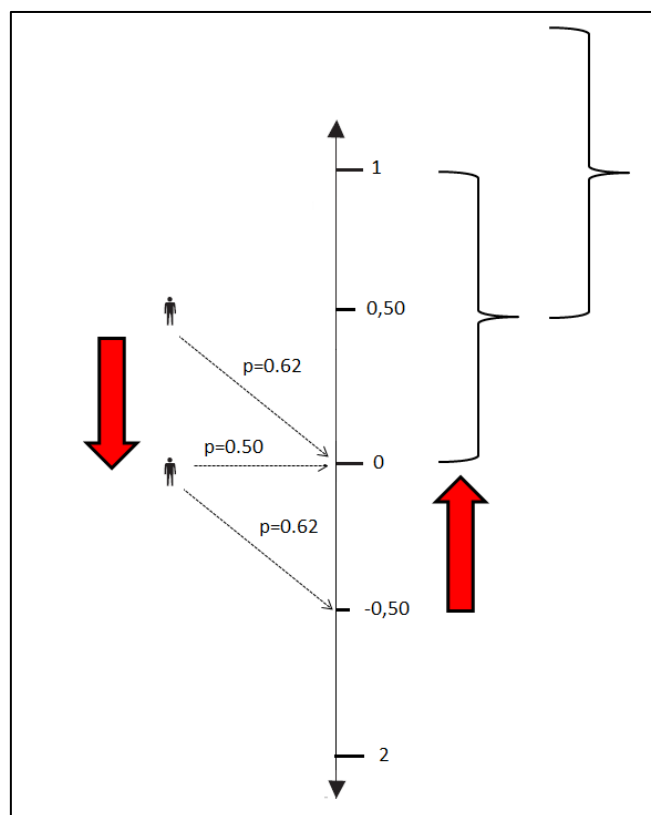
²⁹ Adams, R. & Wu, M. (2010). *PISA 2000 Technical Report*. PISA, OECD Publishing, p 197

Cette probabilité correspond en fait à la probabilité qu'un individu situé au milieu du niveau de compétence a de réussir un item qui se situerait sur le point de césure inférieur du niveau de compétence. Dans le cas présent, cela revient à calculer :

$$p(X_{nj} = 1 | \beta_n = 0.50; \delta_j = 0.00) = \frac{\exp(0.50 - 0)}{1 + \exp(0.50 - 0)} = 0.622459$$

Pour simplifier les représentations graphiques et les explications, il est d'usage de se contenter d'une *Response Probability* qui ne compte que deux décimales, soit 0.62.

Figure 3 : deux procédures d'ajustement pour modifier la *Response Probability*



Comment peut-on modifier l'échelle pour qu'un sujet d'une compétence donnée ait une probabilité de 0,62 de répondre à un item dont la difficulté est égale à sa compétence ? Dans la partie gauche de la figure 3, deux individus sont représentés. Le premier individu, dont la compétence est de 0, a une probabilité de 0,50 de répondre à un item de difficulté de 0.00. Par contre, un autre individu dont la compétence est de 0,50 a une probabilité voisine de 0.62 de réussir ce même item de difficulté 0.00. Si l'on retire la valeur de 0.50 à chaque individu, l'individu dont la compétence initiale est de 0.50 se verra attribuer une compétence de 0.00. Mais sa probabilité de réussir l'item 0.00 n'a pas changé. Il a toujours bien une probabilité voisine de 0,62 de réussir cet item. De même, le second individu a à présent une compétence de -0,50 et il a bien une probabilité de réussir l'item de difficulté -0,50 voisine de 0.62. En diminuant la compétence des sujets et en maintenant constante la difficulté des items, on modifie ainsi la *Response Probability*.

Comme l'illustre la figure 3, on peut aussi modifier l'indice de difficulté des items pour modifier la *Response Probability*. Ainsi, l'individu dont la compétence est de 0.00 a une probabilité voisine de 0.62 de réussir un item dont la difficulté est de -0.50. En augmentant la difficulté des items de 0.50, la *Response Probability* passe ainsi de 0.50 à 0.62. Il convient de noter dans ce cas que les points de césure des niveaux de compétences sont aussi modifiés. Ce changement est graphiquement représenté dans les deux accolades. L'accolade de gauche représente le niveau de compétence initialement défini avec une *Response Probability* de 0.50. L'accolade de droite représente le niveau de compétence avec une *Response Probability* de 0.62.

Dans le cadre de la première évaluation groupée, il a été décidé d'adopter la seconde méthode, à savoir une modification de la *Response Probability* par une élévation des indices de difficulté des items.

Comme susmentionné, une modification des indices de difficulté de 0.50 conduit la *Response Probability* de passer de 0.50 à 0.622459... Pour connaître le changement d'échelle pour obtenir une *Response Probability* parfaitement égale à 0.62, il suffit de calculer

$$\ln\left(\frac{RP}{1-RP}\right) = \left(\frac{0.62}{1-0.62}\right) = 0.4895$$

8.2. Niveaux de compétence et points de césure

Le tableau 1 reprend la longueur des niveaux de compétences, la *Response Probability* et le point de césure le plus bas pour les échelles de lecture et de mathématiques, en deuxième et en sixième années. Les tableaux 2 à 5 présentent, par année d'études, et par discipline, les points de césure initiaux sur l'échelle *Logit*, après modification de la *Response Probability* et enfin après transformation des échelles pour obtenir les échelles internationale PASEC de moyenne 500 et d'écart-type 100.

Tableau 1 : caractéristiques des niveaux de compétences des échelles PASEC.

Cycle	Année	Discipline	Longueur (logit)	Response Probability	Premier point de césure
2014	2	Lecture	1,00	0,62	-2,60
		Mathématiques	1,50	0,68	-2,20
	6	Lecture	0,90	0,62	-1,72
		Mathématiques	0,90	0,62	-1,20
2015	maitres	Lecture	1,30	0,66	-1,90
		Mathématiques	1,10	0,66	-1,40

Tableau 2 : points de césure des niveaux de compétences en Lecture deuxième année

	RP 0,50 Echelle <i>Logit</i>	RP 0,62 Echelle <i>Logit</i>	RP 0,62 Echelle 500,100
Sous le niveau 1	De $-\infty$ à -2,60	De $-\infty$ à -2,11	De $-\infty$ à -399,12
Niveau 1	De -2,60 à -1,60	De -2,11 à -1,11	De 399,12 à 469,54
Niveau 2	De -1,60 à -0,60	De -1,11 à -0,11	De 469,54 à 539,96

Niveau 3	De -0,60 à 0,40	De -0,11 à 0,89	De 539,96 à 610,38
Niveau 4	De 0,40 à +∞	De 0,89 à +∞	De 610,38 à +∞

Tableau 3 : points de césure des niveaux de compétences en Mathématiques deuxième année

	RP 0,50 Echelle <i>Logit</i>	RP 0,68 Echelle <i>Logit</i>	RP 0,68 Echelle 500,100
Sous le niveau 1	De -∞ à -2,20	De -∞ à -1,45	De -∞ à 400,34
Niveau 1	De -2,20 à -0,70	De -1,45 à 0,05	De 400,34 à 489,03
Niveau 2	De -0,70 à 0,80	De 0,05 à 1,55	De 489,03 à 577,73
Niveau 3	De 0,80 à +∞	De 1,55 à +∞	De 577,73 à +∞

Tableau 4 : points de césure des niveaux de compétences en Lecture sixième année

	RP 0,50 Echelle <i>Logit</i>	RP 0,62 Echelle <i>Logit</i>	RP 0,62 Echelle 500,100
Sous le niveau 1	De -∞ à -1,72	De -∞ à -1,23	De -∞ à 365,01
Niveau 1	De -1,72 à 0,82	De -1,23 à -0,33	De 365,01 à 441,69
Niveau 2	De -0,82, à 0,08	De -0,33 à 0,57	De 441,69 à 518,37
Niveau 3	De 0,08 à 0,98	De 0,57 à 1,47	De 518,37 à 595,05
Niveau 4	De 0,98 à +∞	De 1,47 à +∞	De 595,05 à +∞

Tableau 5 : points de césure des niveaux de compétences en Mathématiques sixième année

	RP 0,50 Echelle <i>Logit</i>	RP 0,62 Echelle <i>Logit</i>	RP 0,62 Echelle 500,100
Sous le niveau 1	De -∞ à -1,20	De -∞ à -0,71	De -∞ à 433,28
Niveau 1	De -1,20 à -0,30	De -0,71 à 0,19	De 433,28 à 521,46
Niveau 2	De -0,30 à 0,60	De 0,19 à 1,09	De 521,46 à 609,64
Niveau 3	De 0,60 à +∞	De 1,09 à +∞	De 609,64 à +∞

Tableau 6 : points de césure des niveaux de compétences en Lecture maître

	RP 0,50 Echelle <i>Logit</i>	RP 0,66 Echelle <i>Logit</i>	RP 0,62 Echelle 500,100
Sous le niveau 1	De -∞ à -1,90	De -∞ à -1,237	De -∞ à 290,88
Niveau 1	De -1,90 à -0,60	De -1,237 à 0,063	De 290,88 à 394,22
Niveau 2	De -0,60 à 0,70	De 0,063 à 1,363	De 394,22 à 497,56
Niveau 3	De 0,70 à +∞	De 1,363 à +∞	De 497,56 à +∞

Tableau 7 : points de césure des niveaux de compétences en Mathématiques maître

	RP 0,50 Echelle <i>Logit</i>	RP 0,66 Echelle <i>Logit</i>	RP 0,62 Echelle 500,100
Sous le niveau 1	De $-\infty$ à -1,40	De $-\infty$ à -0,737	De $-\infty$ à 365,09
Niveau 1	De -1,40 à -0,30	De -0,737 à 0,363	De 365,09 à 456,15
Niveau 2	De -0,30 à 0,80	De 0,363 à 1,463	De 456,15 à 547,21
Niveau 3	De 0,80 à $+\infty$	De 1,463 à $+\infty$	De 547,21 à $+\infty$

Transformation linéaire pour passer de l'échelle *Logit* à l'échelle 500/100

Le tableau 7 présente les transformations linéaires pour passer des échelles *Logit* aux échelles 500/100.

Tableau 7: transformations linéaires

	Moyenne	Ecart type
Lecture 2	-0,678	1,420
Mathématiques 2	0,239	1,691
Lecture 6	0,319	1,207
Mathématiques 6	-0,036	1,030
Prof lecture	1,394	1,258
Prof Math	0,893	1,208
Prof lecture didactique	0,087	0,784
Prof math didactique	-0,401	0,942

8.3. Description des échelles et des niveaux de compétences

Compréhension de l'écrit

Tableau 8.6 : Description de l'échelle de compétences PASEC2019 en langue – Début de scolarité

Niveau	Description des compétences
4	<p>Lecteur intermédiaire : vers une langue autonome pour comprendre des phrases et des textes</p> <p>Les élèves ont atteint un niveau de déchiffrage de l'écrit et un niveau de compréhension orale qui leur permettent de comprendre des informations explicites dans des phrases et des textes courts. Ils sont capables de croiser leurs compétences de décodage et leur maîtrise du langage oral pour restituer le sens littéral d'un texte court.</p>
3	<p>Apprenti lecteur : vers le perfectionnement du déchiffrage de l'écrit, des capacités de compréhension orale et de compréhension des mots écrits</p> <p>Les élèves ont perfectionné leurs capacités de compréhension orale et de décodage pour se concentrer sur la compréhension de mots. En compréhension de l'oral, ils sont capables de comprendre des informations explicites dans un texte court dont le vocabulaire est familier. Ils développent progressivement des liens entre langage oral et</p>

	écrit pour améliorer les capacités de décodage et étendre le vocabulaire. En compréhension de l'écrit, les élèves sont capables d'identifier le sens de mots isolés.
2	Lecteur émergent : vers le développement des capacités de déchiffrage de l'écrit et le renforcement des capacités de compréhension orale Les élèves ont amélioré leur niveau de compréhension de l'oral et sont en mesure d'identifier un champ lexical. Ils développent les premiers liens rudimentaires entre le langage oral et écrit, et sont capables de réaliser des tâches basiques de déchiffrage, de reconnaissance et d'identification graphophonologique (lettre, syllabe, graphème, phonème).
1	Le lecteur en éveil : les premiers contacts avec le langage oral et écrit Les élèves sont capables de comprendre des messages oraux très courts (mots isolés) et familiers pour reconnaître des objets familiers. Ils éprouvent de grandes difficultés dans le déchiffrage de l'écrit et l'identification graphophonologique (lettres, syllabes, graphèmes et phonèmes).
Sous le niveau 1	Les élèves qui se situent à ce niveau ne manifestent pas suffisamment les compétences mesurées par ce test en langue de scolarisation. Ils sont en difficulté sur les connaissances et compétences du niveau 1

Tableau 8.7 : Description de l'échelle de compétences PASEC2019 en lecture – Fin de scolarité

Niveau	Description des compétences
4	Les élèves peuvent effectuer un traitement de texte global pour tirer parti de textes narratifs, informatifs et de documents. Sur ces supports, ils sont capables d'associer et d'interpréter plusieurs idées implicites en s'appuyant sur leurs expériences et leurs connaissances. En lisant des textes littéraires, les élèves sont capables d'identifier l'intention de l'auteur et de déterminer le sens implicite d'un récit. En lisant des textes informatifs et des documents, ils mettent en lien des informations et comparent les données pour les exploiter.
3	Les élèves sont capables de combiner deux informations explicites dans un passage de document ou de réaliser des inférences simples dans un texte narratif ou informatif. Ils peuvent extraire des informations implicites de supports écrits en donnant du sens aux connecteurs implicites, aux anaphores ou aux référents. Les élèves localisent des informations explicites dans des textes longs et des documents dont le texte est discontinu.
2	Les élèves améliorent leur capacité de décodage pour comprendre des mots isolés issus de leur vie quotidienne et des phrases isolées. Ils sont également en mesure de localiser des informations explicites dans des textes courts et moyens en prélevant des indices de repérage présents dans le texte et dans les questions. Les élèves parviennent à paraphraser des informations explicites d'un texte.
1	Les élèves ont développé des capacités de décodage et sont capables de les mobiliser pour comprendre des mots isolés issus de leur vie quotidienne ou des phrases isolées très brèves, mais sont en difficulté pour comprendre le sens de textes courts et simples.
Sous le niveau 1	Les élèves qui se situent à ce niveau ne manifestent pas suffisamment les compétences mesurées par ce test en langue d'enseignement. Ils sont en difficulté sur les connaissances et sur les compétences du niveau 1.

Mathématiques

Tableau 8.8 : Description de l'échelle de compétences PASEC2019 en mathématiques – Début de scolarité

Niveau	Compétences des élèves
3	Les élèves maîtrisent la chaîne verbale (compter jusqu'à 60 en deux minutes) et sont capables de lire des chiffres, de comparer des nombres, de compléter des suites de nombres et de réaliser des opérations (additions et soustractions) sur des nombres supérieurs à 50. Ils peuvent raisonner sur des problèmes basiques avec des nombres inférieurs à 20.
2	Les élèves sont capables de lire des chiffres, de comparer des nombres, de compléter des suites logiques et de réaliser des opérations (additions et soustractions) sur des nombres inférieurs à 50. Ils manipulent des concepts de repérage dans l'espace (par ex. devant, sur ...). Ils commencent à développer des aptitudes de raisonnement sur des problèmes basiques avec des nombres inférieurs à 20. Ils identifient aussi la plupart des formes géométriques simples.
1	Les élèves développent progressivement leurs connaissances du langage mathématique : ils commencent à lire les premiers chiffres (inférieurs à 10) et maîtrisent les premières notions de quantité (dénombrement, comparaison) avec des nombres inférieurs à 20. Ils apprécient la taille relative des objets et commencent à identifier de premières formes géométriques simples.
Sous le niveau 1³⁰	Les élèves qui se situent à ce niveau ne manifestent pas suffisamment les compétences mesurées par ce test de mathématiques. Ces élèves sont en difficulté sur les connaissances et compétences de niveau 1.

³⁰ Deux items se situent sous la borne inférieure du niveau 1 : ces deux items relèvent du domaine « géométrie, espace et mesure » et plus spécifiquement de la compétence « estimer des grandeurs ». Il n'est pas possible de définir un niveau de compétence cohérent au départ de ces deux items mais on peut considérer qu'ils sont acquis par les élèves du niveau 1 et des niveaux supérieurs ainsi que par une certaine proportion d'élèves n'atteignant pas le niveau 1.

Tableau 8.9 : Description de l'échelle de compétences PASEC2019 en mathématiques – Fin de scolarité

Niveau	Compétences des élèves
3	Les élèves peuvent effectuer un traitement de texte global pour tirer parti de textes narratifs, informatifs et de documents. Sur ces supports, ils sont capables d'associer et d'interpréter plusieurs idées implicites en s'appuyant sur leurs expériences et leurs connaissances. En lisant des textes littéraires, les élèves sont capables d'identifier l'intention de l'auteur et de déterminer le sens implicite d'un récit. En lisant des textes informatifs et des documents, ils mettent en lien des informations et comparent les données pour les exploiter.
2	Les élèves sont capables de combiner deux informations explicites dans un passage de document ou de réaliser des inférences simples dans un texte narratif ou informatif. Ils peuvent extraire des informations implicites de supports écrits en donnant du sens aux connecteurs implicites, aux anaphores ou aux référents. Les élèves localisent des informations explicites dans des textes longs et des documents dont le texte est discontinu.
1	Les élèves améliorent leur capacité de décodage pour comprendre des mots isolés issus de leur vie quotidienne et des phrases isolées. Ils sont également en mesure de localiser des informations explicites dans des textes courts et moyens en prélevant des indices de repérage présents dans le texte et dans les questions. Les élèves parviennent à paraphraser des informations explicites d'un texte.
Sous le niveau 1	Les élèves ont développé des capacités de décodage et sont capables de les mobiliser pour comprendre des mots isolés issus de leur vie quotidienne ou des phrases isolées très brèves, mais sont en difficulté pour comprendre le sens de textes courts et simples.

Tableau 8.10 : Description de l'échelle de compétences PASEC2019 des enseignants en compréhension de l'écrit

Niveaux	Description des compétences des enseignants
Niveau 3	À ce niveau, les enseignants sont capables de prendre du recul et d'opérer un traitement global sur tout type de textes. Ils réalisent des inférences complexes et parviennent à combiner et interpréter plusieurs idées implicites en s'appuyant sur leurs expériences et leurs connaissances. Les enseignants sont en mesure de se détacher du sens littéral d'un texte pour identifier l'intention de l'auteur, percevoir la dimension humoristique d'un texte (même quand elle est discrète). Ils peuvent tenir compte du contenu d'un texte pour formuler une idée nouvelle pertinente en lien avec les informations lues.
Niveau 2	Les enseignants manifestent leur capacité à utiliser des informations paraphrasées. Ils sont capables de réaliser des inférences simples dans tout type de texte. Ils parviennent également à percevoir le réseau anaphorique d'un texte littéraire. Les enseignants sont capables de combiner des informations présentes dans différentes parties d'un texte.
Niveau 1	Les enseignants sont en mesure de localiser des informations explicites dans des textes moyens ou longs en utilisant des indices de repérage présents dans le texte et dans les questions. Ils mobilisent cette compétence sur des textes narratifs et informatifs. Les enseignants parviennent à repérer quelques paraphrases élémentaires dans un texte.
Sous le niveau 1	Les enseignants qui se situent à ce niveau ne manifestent pas suffisamment les compétences mesurées par ce test en compréhension de l'écrit. Ils sont en difficulté sur les connaissances et compétences du niveau 1.

Tableau 8.11 : Description de l'échelle de compétences PASEC2019 des enseignants en mathématiques

Niveaux	Description des compétences des enseignants
Niveau 3	Les enseignants situés à ce niveau font montre de capacités à résoudre des problèmes complexes, impliquant plusieurs étapes de résolution et nécessitant la mise en œuvre d'un raisonnement s'appuyant sur une analyse approfondie de la situation et pouvant impliquer de manipuler des inconnues (par exemple dans les problèmes de partages inégaux). Pour résoudre la plupart des tâches de ce niveau, une vigilance cognitive est nécessaire pour éviter les erreurs courantes et inhiber les conceptions erronées (par exemple penser que l'aire et le périmètre varient dans le même sens). L'expertise caractéristique de ce niveau traduit aussi une compréhension approfondie des concepts (par exemple appréhender la relativité du tout face au concept de fraction).
Niveau 2	Les enseignants situés à ce niveau peuvent résoudre de nombreux problèmes de proportionnalité directe, ainsi que des problèmes complexes, impliquant plusieurs étapes de résolution et nécessitant la mise en œuvre d'une démarche organisée de façon séquentielle. Plusieurs tâches de ce niveau impliquent des conversions d'unités, intégrées ou non dans des situations-problèmes. Certaines tâches font appel à des connaissances factuelles portant sur des objets mathématiques diversifiés (par exemple, les formules de calcul d'aire d'un solide, une propriété des triangles, ou encore le nom d'un triangle particulier...).
Niveau 1	Les enseignants situés à ce niveau témoignent de connaissances factuelles et d'une maîtrise de procédures de base qu'ils mettent en œuvre dans les tâches d'application directe. Leurs compétences se situent essentiellement dans le domaine des nombres (par exemple la connaissance de l'écriture décimale, la comparaison de fractions, la notion de pourcentage ...) et des opérations (capacités à résoudre des opérations impliquant des nombres entiers, des nombres décimaux et des fractions). Ils témoignent aussi de quelques connaissances dans le domaine des solides et figures (par exemple, identifier des figures ou des solides, repérer une diagonale ou un axe de symétrie ...) ainsi qu'en grandeurs et mesures (ex. calculer le périmètre d'un triangle). Peu de tâches de résolution de problèmes se situent à ce niveau.
Sous le niveau 1	Les enseignants qui se situent à ce niveau ne manifestent pas suffisamment les connaissances et les compétences mesurées par ce test. Ces enseignants sont en difficulté sur les connaissances et compétences de niveau 1.

L'évaluation PASEC2019 a permis de collecter diverses informations sur les caractéristiques des élèves, leur environnement familial, leurs perceptions et l'environnement scolaire (enseignants, classes, directeurs et écoles).

La plupart de ces informations se présentent sous forme de questions ou de variables directement exploitables (comme l'âge et le genre de l'élève). Rappelons cependant que des recodages simples peuvent être nécessaires afin de faciliter les analyses et les interprétations (par exemple : recodage de la variable genre de l'élève en 1 pour les filles et 0 pour les garçons).

Cependant, plusieurs variables sont conçues afin d'être agrégées pour mesurer un trait latent (par exemple la perception, l'attitude...) qu'il serait difficile de mesurer directement avec une seule variable. Comme pour les compétences des élèves mesurées par une analyse synthétique d'items, ces questions contextuelles sont agrégées et mises à l'échelle à l'aide de la théorie de réponse à l'item (IRT) pour créer des indices contextuels.

Nous présentons dans ce chapitre la méthode de création des indices au niveau international pour des données contextuelles.

Liste des indices contextuels analysés

Sur la base du cadre de référence des questionnaires PASEC2019, les données sont collectées en vue de créer les indices suivants :

Au niveau élèves³¹ (6^e année) :

- Indice du statut socioéconomique

Au niveau directeurs/écoles

- Indice d'infrastructures de l'école
- Indice d'implication de la communauté
- Indice d'aménagement du territoire

Certains indices³² n'ont pu être créés à cause d'une cohérence globale très faible dans la majorité des pays. La liste des items composant ces indices se trouve dans le manuel d'exploitation des données.

8.4. Procédure de création des indices contextuels

La démarche qui a été suivie est celle de la théorie de réponse à l'item (IRT). Pour les indices composés d'items³³ entièrement dichotomiques (par exemple : indice des avantages sociaux de l'enseignant, indice d'aménagement du territoire), le modèle de Rasch à un paramètre a été utilisé.

La spécification du modèle Rasch se présente comme suit :

$$P_i(\theta_n) = \frac{\exp(\theta_n - \delta_i)}{1 + \exp(\theta_n - \delta_i)}$$

Où :

$P_i(\theta_n)$ est la probabilité qu'un individu n réponde « Oui » à l'item i ;

θ_n est l'estimation du trait latent de l'individu n ;

δ_i est l'estimation de la position de l'item i sur l'échelle.

Pour les indices composés d'items du type « Likert » ou ceux présentant plus de deux modalités, le modèle précédent peut être substitué à un modèle de crédit partiel qui se présente comme suit :

³¹ L'enquête PASEC2019 n'a pas collecté ces données auprès des élèves de 2e année. En effet, il n'est pas possible de garantir la fiabilité de telles données pour ce niveau d'études.

Au niveau maîtres/classes :

- Indice de gestion de la discipline

Au niveau directeurs/écoles :

- Indice de la qualité de la gestion administrative
- Indice de la qualité de la gestion pédagogique

³³ Variables primaires.

$$P_{x_i}(\theta_n) = \frac{\exp \sum_{k=0}^x (\theta_n - \delta_i + \tau_{ik})}{\sum_{h=0}^{m_i} \exp \sum_{k=0}^h (\theta_n - \delta_i + \tau_{ik})}$$

Avec $x_i = 0, 1, \dots, m_i$

Où :

$P_{x_i}(\theta_n)$ est la probabilité qu'un individu n ait un crédit partiel de x à l'item i (x variant entre 0 et m_i) ;

θ_n est l'estimation du trait latent de l'individu n ;

δ_i est l'estimation de la position de l'item i sur l'échelle ;

τ_{ij} est un paramètre relatif au « pas³⁴ » entre deux crédits partiels successifs.

Les principales étapes suivies pour la création de l'ensemble des indices sont les suivantes :

- 1) recodage des items pour une utilisation appropriée du modèle IRT ;
- 2) première analyse IRT des variables de chaque indice par pays : stabilisation de la liste des indices sur la base de la cohérence interne globale (alpha de Cronbach supérieur ou égal à 0,60) ;
- 3) analyse des paramètres psychométriques des items des indices par pays : identification des items défectueux sur la base de la discrimination et de l'ajustement au modèle IRT pour chaque pays ;
- 4) synthèse des analyses nationales : identification du fonctionnement différentiel des items à travers les pays ;
- 5) estimation internationale des paramètres d'item ;
- 6) estimation des scores pour chaque indice.

Différents programmes ont été développés pour le traitement et l'analyse des données lors de ce processus de création des indices. Les logiciels utilisés sont Stata (pour le traitement) et ACER ConQuest® pour les analyses psychométriques.

Recodage des données brutes

Les données ont été recodées pour les besoins de l'analyse IRT. Il s'agit principalement de :

- recoder les items dichotomiques en 0 ou 1, le code 1 représentant l'attribution de l'élément en question ;
- recoder les items catégoriels (de type Likert par exemple) ou les items nécessitant la fusion de plusieurs variables à crédit partiel (codes 0, 1, 2...);
- recoder les données manquantes par le code 7 signifiant « non-réponse par design ».

Encadré 9.1 : Exemple de code Stata pour le recodage et le calcul des alphas de Cronbach

³⁴ Le PASEC a utilisé des « pas » non constants, c'est-à-dire en fonction de chaque item.

```

/* Indice de perception de soi des élèves */

foreach var of varlist qe630a qe630b qe630c {
    recode `var' (1=0) (2=1) (3=2) (4=3)
}

foreach var of varlist qe630d qe630e qe630f qe630g {
    recode `var' (4=0) (3=1) /*(2=2)*/ (1=3)
}

alpha qe630a-qe630g, std item
bysort pays: alpha qe630a-qe630g, std item

```

Cohérence interne des indices

L'indicateur de cohérence interne (ou de fiabilité) a été calculé pour chaque indice et pour chaque pays. Selon les procédures PASEC, le seuil minimum de cohérence interne (alpha de Cronbach) dans le cadre de l'analyse de données contextuelles est de 0,60.

L'alpha de Cronbach se définit comme suit :

$$\alpha = \frac{k}{k-1} \left(1 - \frac{\sum_{i=1}^k \sigma_{Y_i}^2}{\sigma_X^2} \right)$$

Où :

- k est le nombre d'items ;
- σ_X^2 est la variance du score total ;
- $\sigma_{Y_i}^2$ est la variance de l'item i .

Analyse psychométrique des paramètres d'item des indices par pays

La qualité psychométrique des indices est appréciée en premier lieu pour chaque pays. Pour ce faire, deux indicateurs sont utilisés : la discrimination et l'ajustement. L'indice de discrimination mesure la corrélation entre la réponse à l'item et les réponses à l'ensemble des items de l'indice. Quant à l'indice d'ajustement, il mesure la qualité d'ajustement du modèle (logistique) aux données observées.

Pour chaque indice, les items défectueux sont identifiés sur la base des règles suivantes de discrimination et d'ajustement au modèle :

- la discrimination³⁵ (ou la corrélation bisériale de point) de l'item est inférieure à 0,20 ;
- l'indice d'ajustement pondéré est inférieur à 0,75 ou supérieur à 1,25, pour autant que la statistique T soit supérieure en valeur absolue à 1,96.

³⁵ Voir la section « Analyse et calibration des items » au chapitre 7.

Encadré 9.2 : Exemple de code ConQuest pour l'estimation des caractéristiques psychométriques (discrimination et ajustement) des items

```

/*      Indice SES : (24 items; codes 0,1,2,3)      */
let path=C:\PASEC\Analyse_IRT_QUESTIONNAIRES_2015;
Dofor itere=1 2 3 4 5 6 7 8 9 10 11;

set update=yes,warnings=no,iterlimit=5000,constraints=items;
datafile %path%\P%itere%\ELEVES_6A_P%itere%.txt;
format pays 8-9 id 17-20 response 50-73;
labels << %path%\Fichiers_LAB\el_ses.lab;
codes 0,1,2,3;
model item + item*step;
estimate ! method=quadrature, fit=yes, conv=0.001, stderr=Empirical;
show >> %path%\P%itere%\eleves\ELEVES_P%itere%_ses1.shw;
show parameters ! table=2,filetype=excel >>
%path%\P%itere%\eleves\ELEVES_P%itere%_ses1.xls;
itanal >> %path%\P%itere%\eleves\ELEVES_P%itere%_ses1.itn;
itanal! estimates=latent,filetype=excel,format=summary >>
%path%\P%itere%\eleves\ELEVES_P%itere%_ses1_sum.xls;
enddo;
quit;

```

Quand un item dysfonctionne en termes de discrimination dans la majorité des pays, c'est-à-dire dans au moins cinq pays, l'item est supprimé définitivement pour tous les pays. Un item qui n'est pas supprimé au niveau international ne l'est pas non plus au niveau d'un pays si le dysfonctionnement en termes de discrimination provient uniquement d'une faible variance. En effet, une faible discrimination (discrimination positive mais inférieure à 0,20) peut traduire une faible variance des données relativement à l'item en question.

La liste des items supprimés pour tous les pays pour cause de dysfonctionnement par rapport à certains indices se présente comme suit :

Synthèse des analyses nationales

Après l'identification des dysfonctionnements des items pour chaque pays, le fonctionnement différentiel entre pays a été vérifié. Il s'agissait de vérifier si un item qui présente un bon fonctionnement au sein de chaque pays peut présenter des indices statistiques qui diffèrent substantiellement d'un pays à l'autre ; autrement dit, si pour un pays donné un item mesure un trait latent qui diffère plus ou moins de celui mesuré par le même item dans d'autres pays.

Un item présente un fonctionnement différentiel entre pays en termes d'indice de difficulté (*item by country interaction*)³⁶ si l'écart entre l'indice de difficulté de l'item au sein d'un pays et la moyenne des indices nationaux est supérieur en valeur absolue à 0,50, pour autant que sa valeur standardisée (l'écart rapporté à l'erreur type) en valeur absolue soit supérieure à 1,96.

Estimation internationale des paramètres d'item

À la suite des analyses nationales, l'estimation internationale des paramètres d'item a permis d'apprécier la qualité psychométrique des indices au niveau international. Afin de ne pas avantager ou désavantager les pays³⁷, les items présentant des discriminations faibles au niveau national sont maintenus dans ces pays lors du calibrage international. Un poids normalisé³⁸ a été utilisé afin que chaque pays puisse contribuer de manière identique à l'estimation des paramètres finaux des items. Pour le Cameroun, le poids attribuable au pays a été réparti proportionnellement aux poids des deux sous-systèmes éducatifs (soit 30 % pour le système anglophone et 70 % pour le système francophone).

Les mêmes règles ont été utilisées pour mesurer le bon fonctionnement des items en termes de discrimination et d'ajustement.

Pour chacun des dix indices, tous les items présentent une bonne discrimination. L'ajustement faible qui est observé, quelques fois, pour certains items s'explique par la discrimination élevée car inversement corrélé à cette dernière.

³⁶ Nous nous limitons dans ce rapport à analyser l'« item by country interaction », même s'il est possible également de vérifier le fonctionnement différentiel en termes de discrimination ou d'ajustement dans les conditions suivantes :

- l'indice de discrimination de l'item au sein d'un pays s'écarte d'au moins 0,10 de la moyenne des valeurs nationales ;
- l'indice d'ajustement pondéré de l'item au sein d'un pays s'écarte d'au moins 0,10 par rapport à la moyenne des valeurs nationales pour le même item.

³⁷ Quand un item est faiblement possédé (par exemple par 3 % de l'échantillon) dans un pays, la discrimination pourrait s'afficher faible et supprimer un tel item avantagerait ce pays par rapport à l'indice. Idem pour un item fortement possédé (par exemple 99 % de possession) : le supprimer désavantagerait le pays concerné par rapport à l'indice.

³⁸ Par exemple, pour les indices de niveau élèves : $POIDS_NORMALISE = 1000 * POIDS / (\sum POIDS)$.

Encadré 9.3 : Code ConQuest d'estimation internationale des caractéristiques psychométriques (discrimination et ajustement) des items

```

/* Indice SES : (22 items; codes 0,1,2,3) */
let path=C:\PASEC\Analyse_IRT_QUESTIONSNAIRES_2015;
let name=ELEVES_6A;

/* estimation des paramètres internationaux qui seront utilisés pour
estimer les scores */
set update=yes,warnings=no,iterlimit=5000,constraints=items;
datafile %path%\international\eleves\%name%_step3.txt;
format pays 8-9 id 16-20 response 24-45 poids 63-72;
labels << %path%\Fichiers_LAB\INT\el_ses_int.lab;
codes 0,1,2,3;
caseweight poids;
model item + item*step - pays;
estimate ! method=quadrature, fit=yes, conv=0.001, stderr=Empirical;
export parameters >> %path%\international\eleves\%name%_ses_par1.par;
show >> %path%\international\eleves\%name%_ses1.shw;
show parameters ! table=2,filetype=excel >>
%path%\international\eleves\%name%_ses1.xls;
reset;

/* estimation des paramètres internationaux pour l'analyse d'item */
set update=yes,warnings=no,iterlimit=5000,constraints=items;
datafile %path%\international\eleves\%name%_step3.txt;
format pays 8-9 id 16-20 response 24-45 poids 63-72;
labels << %path%\Fichiers_LAB\INT\el_ses_int.lab;
codes 0,1,2,3;
caseweight poids;
model item + item*step;
estimate ! method=quadrature, fit=yes, conv=0.001, stderr=Empirical;
itanal >> %path%\international\eleves\%name%_ses2.itn;
itanal! estimates=latent,filetype=excel,format=summary >>
%path%\international\eleves\%name%_ses2_sum.xls;
show parameters ! table=2,filetype=excel >>
%path%\international\eleves\%name%_ses2.xls;
put >> %path%\international\eleves\%name%_ses.cqs;
quit;

```

Estimation des scores

Sur la base des paramètres internationaux des items, le « *Weighted Likelihood Estimation* » (WLE) (Warm, 1989)³⁹ a été utilisé pour estimer les scores pour les élèves, les maîtres et les directeurs d'école. Ces scores ont ensuite été standardisés en une moyenne internationale de 50 et un écart-type de 10 pour chacun des indices.

³⁹ Warm, T. A. (1989). Weighted Likelihood Estimation of Ability in Item Response Theory. *Psychometrika*, 54(3), 427-450.

Encadré 9.4 : Exemple de code ConQuest pour l'estimation des scores WLE des élèves pour l'indice socioéconomique

```

/*   Indice SES : (22 items; codes 0,1,2,3)   */
let path=C:\PASEC\Analyse_IRT_QUESTIONNAIRES_2015\IRT;
let name=ELEVES_6A;

set update=yes,warnings=no,iterlimit=5000,constraints=items;
datafile %path%\international\eleves\%name%_step3.txt;
format pays 8-9 pid 16-20 response 24-45 poids 63-72;
labels << %path%\Fichiers_LAB\INT\el_ses_int.lab;
codes 0,1,2,3;
model item + item*step;
import anchor_parameters <<
%path%\international\eleves\%name%_ses_par2.par;
estimate ! method=quadrature, fit=yes, nodes=30, conv=0.001,
stderr=Empirical;
show cases ! estimate=wle,filetype=excel >>
%path%\international\eleves\%name%_ses_wle.xls;
show >> %path%\international\eleves\%name%_ses_wle.shw;

```

La formule de standardisation pour l'obtention d'une moyenne de 50 et d'un écart-type de 10 est la suivante :

$$Score_{std} = \frac{Score_{wle} - \overline{Score_{wle}}}{\sigma_{Score_{wle}}} \times 10 + 50$$

Où :

$Score_{wle}$ est le score WLE obtenu à partir des analyses IRT ;
 $\overline{Score_{wle}}$ et $\sigma_{Score_{wle}}$ sont respectivement la moyenne et l'écart-type du score WLE.

La distribution pondérée des scores (standardisés) des indices se présente comme suit :

10. CALCUL DE L'ERREUR TYPE

La nouvelle méthodologie d'analyse quantitative du PASEC a connu deux changements majeurs : l'utilisation des réplifications et le recours aux valeurs plausibles. Les réplifications permettent le calcul de l'erreur d'échantillonnage et les valeurs plausibles, au-delà des avantages connus des modèles de réponse à l'item, permettent le calcul de l'erreur de mesure.

10.1 Sources d'erreur : Erreur d'échantillonnage et erreur de mesure

Erreur d'échantillonnage

Le chapitre 6 a montré la procédure de génération des poids issus de réplifications. Ces poids de réplifications permettent ensuite de calculer la variance d'échantillonnage d'une statistique telle que la moyenne, l'écart-type, le coefficient de régression, les quartiles, etc.

La variance d'échantillonnage d'une statistique est calculée en comparant l'estimation de cette statistique pour chaque poids de réplifications h (rwt_h) à l'estimation de la même statistique pour l'échantillon entier (avec le poids final de l'élève, rwt_0).

De façon technique, la procédure de calcul pour l'estimation d'une statistique telle que la moyenne ou les quartiles se fait comme suit :

- La statistique (moyenne, quartile, etc.) est calculée pour l'échantillon complet en pondérant les données avec le poids final élève (rwt_0) ;
- La statistique est calculée pour chacune des 90 répliques en 5^e/6^e année (ou des 45 répliques en 2^e année) en utilisant les poids de réplifications (rwt_h).

La variance d'échantillonnage est ensuite calculée comme suit :

$$\sigma_{(\hat{\theta})}^2 = \sum_{h=1}^R (\hat{\theta}_{(h)} - \hat{\theta})^2$$

Où :

R correspond au nombre de réplifications (90 en 5^e/6^e année et 45 en 2^e année) ;

$\hat{\theta}$ correspond à la statistique calculée sur l'échantillon complet avec le poids final rwt_0 ;

$\hat{\theta}_{(h)}$ correspond à la statistique calculée avec le poids de réplifications h rwt_h .

Erreur de mesure

Depuis sa création, le PASEC a utilisé successivement deux méthodes pour estimer la performance des élèves : la méthode basée sur la théorie du score vrai et la méthode qui s'appuie sur les modèles de réponse à l'item.

La théorie du score vrai permet l'estimation de la performance de l'élève comme le pourcentage de bonnes réponses données par un élève à un test qui lui a été soumis. Ce type de score a été utilisé par le PASEC entre 1991 et 2011.

Les modèles de réponse à l'item ont été utilisés pour la première fois par le PASEC en 2012. Cette méthode repose sur le modèle probabiliste de Rasch (modèle à un paramètre) selon lequel la réussite

d'un élève à un item dépend à la fois du niveau de difficulté de l'item et du niveau de compétence de l'élève.

Les modèles de réponse à l'item ont d'abord été utilisés pour estimer des compétences ponctuelles (WLE⁴⁰) des élèves (en 2012 et 2013), puis pour estimer des valeurs plausibles (à partir de 2014). Les analyses de l'évaluation PASEC2019 reposent sur cinq valeurs plausibles pour chaque élève. Les valeurs plausibles, comparativement au score classique ou aux WLE, ont l'avantage de fournir des estimations non biaisées des éléments de la distribution des scores.

À partir des valeurs plausibles, il est possible de calculer la performance moyenne d'un pays ou d'un groupe d'élèves en déterminant la moyenne des moyennes de chacune des cinq valeurs plausibles. La performance moyenne globale est calculée comme suit :

$$\hat{\mu} = \frac{1}{P} \sum_{i=1}^P \mu_i$$

Où :

μ_i correspond à la moyenne de la performance calculée à partir de la valeur plausible i ;

P correspond au nombre de valeurs plausibles.

L'utilisation des valeurs plausibles permet le calcul de l'erreur de mesure associée à l'estimation de la compétence de l'élève. Cette erreur de mesure est obtenue à partir de la formule suivante :

$$\sigma_{(PV)}^2 = \frac{1}{P-1} \sum_{i=1}^P (\mu_i - \hat{\mu})^2$$

Où :

P correspond au nombre de valeurs plausibles ;

μ_i correspond à la moyenne de la performance calculée à partir de la valeur plausible i ;

$\hat{\mu}$ correspond à la performance moyenne globale.

Pour chaque valeur plausible, une variance d'échantillonnage est calculée en utilisant la formule présentée à la section 10.1.1. Avec P ($P > 1$) valeurs plausibles, l'erreur d'échantillonnage se calcule de la manière suivante :

$$\sigma_{(\hat{\mu})}^2 = \frac{1}{P} \sum_{i=1}^P \sigma_{(\hat{\mu}_i)}^2$$

Les notations précédentes sont conservées.

Erreur totale

L'erreur totale associée à une estimation est égale à la somme des erreurs de mesure et d'échantillonnage. Dans le cas de l'utilisation de P valeurs plausibles et de R répliques, la valeur de cette erreur est obtenue en appliquant la formule suivante :

⁴⁰ *Weighted Likelihood Estimates.*

$$\sigma^2 = \frac{1}{R-1} \sum_{i=1}^R (\mu_i - \hat{\mu})^2 + \frac{1}{P} \sum_{i=1}^P \sigma_{(\hat{\mu}_i)}^2$$

Les notations précédentes sont conservées.

En l'absence de valeurs plausibles, l'erreur totale associée à une estimation est réduite à l'erreur d'échantillonnage.

10.2 Nature des échantillons et calcul de l'erreur type d'une différence

La nature des échantillons en comparaison influence la forme de l'erreur type de l'estimation d'une différence. On distinguera deux cas : le cas d'échantillons indépendants et le cas d'échantillons dépendants. On distinguera aussi les cas avec valeurs plausibles et sans valeurs plausibles.

Échantillons indépendants

Supposons que l'on veuille comparer les performances des élèves d'un pays X à celles des élèves d'un pays Y . Les deux échantillons sont indépendants. L'erreur type d'une différence, dans le cadre de deux échantillons indépendants, est obtenue en appliquant la formule suivante :

$$\sigma_{(X-Y)}^2 = \sqrt{\sigma_X^2 + \sigma_Y^2}$$

Où :

σ_X correspond à l'erreur type associée à l'estimation des scores dans le pays X ;

σ_Y correspond à l'erreur type associée à l'estimation des scores dans le pays Y .

Échantillons dépendants

Lorsque les échantillons ne sont pas indépendants (par exemple, on veut comparer la moyenne de deux variables pour les mêmes élèves), la formule précédente n'est plus valide puisqu'il y a corrélation entre les distributions d'échantillonnage des deux variables.

Sans valeurs plausibles sur les variables à comparer, l'erreur type de la différence est estimée à partir de la formule suivante :

$$\sigma_{(X-Y)} = \sqrt{\left[\sum_{r=1}^R ((\bar{X} - \bar{Y}) - (X^r - Y^r))^2 \right]}$$

Où :

\bar{X} correspond à la valeur moyenne de la variable X sur l'échantillon pondéré par le poids original ;

\bar{Y} correspond à la valeur moyenne de la variable Y sur l'échantillon pondéré par le poids original ;

X^r correspond à la valeur moyenne de la variable X sur l'échantillon pondéré par le poids répliqué r ;

Y^r correspond à la valeur moyenne de la variable Y sur l'échantillon pondéré par le poids répliqué r ;
 R correspond au nombre de réplifications.

Avec des valeurs plausibles, la formule est la suivante :

$$\sigma_{(X-Y)} = \sqrt{\left[\frac{1}{P} \sum_{p=1}^P \left(\sum_{r=1}^R ((X_p^r - Y_p^r) - (X_p - Y_p))^2 \right) \right] + \left[\left(1 + \frac{1}{P} \right) \frac{\sum_{p=1}^P ((X_p - Y_p) - (\bar{X} - \bar{Y}))^2}{P-1} \right]}$$

Où :

X_p^r correspond à la moyenne de la variable X sur la valeur plausible p pour l'échantillon répliqué r ;

Y_p^r correspond à la moyenne de la variable Y sur la valeur plausible p pour l'échantillon répliqué r ;

X_p correspond à la moyenne de la variable X sur la valeur plausible p pour l'échantillon initial ;

Y_p correspond à la moyenne de la variable Y sur la valeur plausible p pour l'échantillon initial ;

\bar{X} correspond à la moyenne estimée de la variable X prenant en compte toutes les valeurs plausibles ;

\bar{Y} correspond à la moyenne estimée de la variable Y prenant en compte toutes les valeurs plausibles ;

R correspond au nombre de réplifications ;

P correspond au nombre de valeurs plausibles.

Dans l'optique de vérifier si un pays remplit certains standards internationaux, on peut être intéressé à évaluer la différence entre la statistique du pays et celle au niveau international. Pour le calcul d'une telle différence, il est à noter qu'une covariance existe aussi entre la statistique du pays X et celle au niveau internationale. La formule de l'erreur type de la différence entre une statistique nationale et une statistique internationale est :

$$\sigma_{(X-I)}^2 = \sqrt{\sigma_X^2 + \frac{\left[\left(\sum_{n=1}^N w_n \right) - w_I \right]^2 - w_I^2}{\left(\sum_{n=1}^N w_n \right)^2} \sigma_I^2}$$

Où :

σ_X correspond à l'erreur associée à l'estimation des scores dans le pays X ;

σ_I correspond à l'erreur associée à l'estimation des scores au niveau international ;

w_n correspond à la somme des poids des élèves dans le pays n ;

w_I correspond à la somme des poids des élèves au niveau international ;

N correspond au nombre total de pays.

10.3 Quelques exemples de calcul de statistiques

Les formules présentées dans ce chapitre peuvent être utilisées pour estimer l'erreur type de n'importe quelle statistique. Les utilisateurs des données du PASEC2019 peuvent par exemple calculer l'erreur type pour :

- une moyenne ;
- un écart-type ;
- un coefficient de corrélation ;
- un coefficient de régression ;
- un coefficient de détermination ;
- etc.

Le PASEC a développé des macros Stata qui permettent d'effectuer ces calculs. Le manuel d'exploitation des données présente comment ces macros peuvent répondre aux besoins des utilisateurs.
