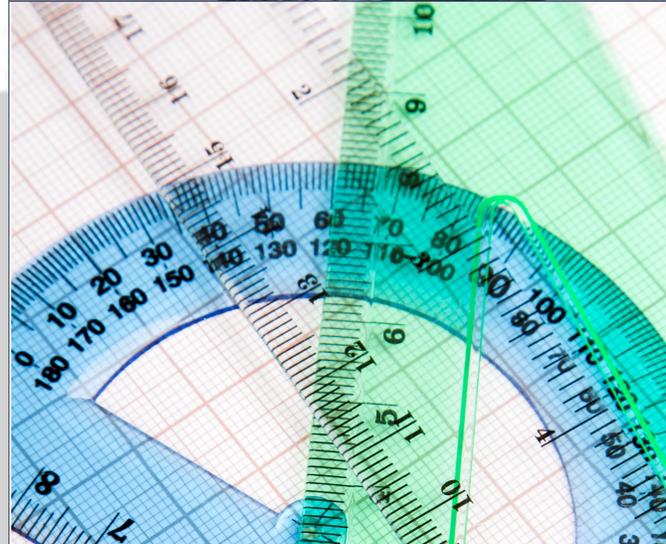


PPCE 2016

Rapport technique



Programme pancanadien d'évaluation

PPCE 2016

Rapport technique

Auteures

M^{me} Kathryn O'Grady, Conseil des ministres de l'Éducation (Canada)

M^{me} Karen Fung, Conseil des ministres de l'Éducation (Canada)



cme^c

Conseil des
ministres
de l'Éducation
(Canada)

Council of
Ministers
of Education,
Canada

Le Conseil des ministres de l'Éducation (Canada) [CMEC] a été créé en 1967 par les ministres responsables de l'éducation désireux de se donner un forum où ils pourraient discuter d'enjeux communs, entreprendre des initiatives sur l'éducation et promouvoir les intérêts des provinces et des territoires auprès des organisations pancanadiennes du secteur de l'éducation, du gouvernement fédéral, des gouvernements étrangers et des organisations internationales. Le CMEC est le porte-parole pancanadien de l'éducation au Canada et, par son entremise, les provinces et territoires travaillent ensemble à l'atteinte d'objectifs couvrant un large éventail d'activités, y compris dans le domaine de l'éducation de la petite enfance et des niveaux primaire, secondaire et postsecondaire.

Par l'entremise du Secrétariat du CMEC, le Conseil agit à titre d'organisation au sein de laquelle les ministères de l'Éducation entreprennent conjointement activités, projets et initiatives, dans des domaines qui intéressent toutes les provinces et tous les territoires¹. L'une de ces activités consiste à élaborer et à administrer des évaluations pancanadiennes basées sur des recherches d'actualité et des pratiques exemplaires en matière d'évaluation du rendement des élèves dans les matières de base.

Remerciements

Le Conseil des ministres de l'Éducation (Canada) remercie les élèves, le personnel enseignant et le personnel administratif de leur participation, sans laquelle le Programme pancanadien d'évaluation n'aurait pu être un succès. C'est grâce à la qualité de votre engagement que cette étude a été possible. Nous vous sommes profondément reconnaissants de votre contribution, qui a permis de broser un tableau pancanadien des politiques et des pratiques d'enseignement des mathématiques, des sciences et de la lecture aux élèves de 8^e année/2^e secondaire².

Conseil des ministres de l'Éducation (Canada)
95, avenue St. Clair Ouest, bureau 1106
Toronto (Ontario) M4V 1N6
Téléphone : 416 962-8100
Télécopieur : 416 962-2800
Courriel : cmec@cmec.ca

© 2019 Conseil des ministres de l'Éducation (Canada)

This report is also available in English.

¹ Ce rapport englobe l'ensemble des provinces et territoires qui participent à l'évaluation.

² Au Québec, ce sont les élèves de 2^e secondaire qui participent au PPCE; dans le reste du Canada, ce sont les élèves de 8^e année.

Table des matières

Chapitre 1. Vue d'ensemble du Programme pancanadien d'évaluation	8
Contexte.....	8
Évaluation pancanadienne.....	8
Participation.....	9
Durée de l'évaluation.....	9
Le PPCE dans les deux langues officielles	10
Chapitre 2. Conception et élaboration de l'évaluation.....	11
Cycle d'évaluation du PPCE.....	11
Présentation des résultats du PPCE au fil du temps.....	12
Mise à jour du cadre d'évaluation	12
Conception de l'évaluation	13
Groupes de travail.....	14
Mise à jour du cadre d'évaluation en lecture.....	14
Élaboration des items	15
Traduction et examen des items	16
Vérification des items d'évaluation et des guides de codage	16
Traduction et comparaison des items en français et en anglais	17
Révisions linguistiques et typographiques	17
Révisions techniques.....	17
Révisions psychométriques	18
Groupe consultatif sur l'équité des tests.....	19
LIGNES DIRECTRICES À L'INTENTION DU GROUPE CONSULTATIF SUR L'ÉQUITÉ DES TESTS ...	19
Approbation des items par les provinces	21
Chapitre 3. Élaboration des questionnaires contextuels	22
Mise à jour du cadre des questionnaires.....	22
Identité sexuelle.....	23
Différences entre les sexes	23
Confidentialité.....	24
Chapitre 4. Procédures d'échantillonnage.....	25
Plan d'échantillonnage.....	25
Stratification.....	27
Taille des échantillons.....	29
Échantillonnage au premier degré – échantillonnage des écoles	30
Bases de données sur les écoles	30
Sélection des écoles	30
Exclusion d'écoles	30
Échantillonnage au deuxième degré – échantillonnage des élèves.....	31
Pondération de l'échantillon	33

Chapitre 5. Mise à l'essai	34
Groupe de travail sur l'examen des items — Mise à l'essai	34
Livrets d'évaluation.....	34
Séance de correction	34
Saisie des données.....	34
Analyse des données.....	35
Groupe de travail sur la sélection des items – Étude principale	36
Révision du cadre d'évaluation.....	36
Chapitre 6. Étude principale	37
Livrets d'évaluation.....	37
Réexamen du matériel d'évaluation.....	37
Impression des livrets d'évaluation	37
Vérification des documents	38
Lettre envoyée aux parents/tutrices ou tuteurs des élèves.....	38
Procédures administratives	38
Lieu de l'évaluation	39
Déroulement de l'évaluation	39
Élèves ayant des besoins spéciaux.....	39
Questionnaires pour la direction de l'école et pour le personnel enseignant	40
Participation des élèves à l'évaluation et exemptions	41
Organisation d'une séance de rattrapage	42
Renvoi du matériel d'évaluation.....	43
Séance de correction	43
Mise en liasse des livrets	43
Feuilles de correction.....	44
Guide de la correctrice ou du correcteur	44
Guide de codage	44
Chefs de correction	44
Chefs de table	45
Formation des correctrices et correcteurs	45
Fidélité des corrections.....	46
Examens de fidélité.....	46
Fiabilité entre membres du groupe de correction (double correction)	48
Fiabilité des tendances	49
Rapports et commentaires	49
Rapport de la coordonnatrice ou du coordonnateur de la province	49
Rapport de la coordonnatrice ou du coordonnateur de l'école.....	51
Formulaire de commentaires de la correctrice ou du correcteur	53
Chapitre 7. Établir une norme de rendement	54
Séances d'établissement des normes.....	54
Sélection d'un groupe de spécialistes.....	54
Versions préliminaires des descripteurs des niveaux de rendement.....	55

Sécurité des documents.....	55
Méthode du signet.....	55
Procédure d'établissement des normes.....	57
Descripteurs des niveaux de rendement.....	59
Chapitre 8. Traitement des données du PPCE.....	61
Collecte des données.....	61
Saisie des données.....	61
Contrôle de la qualité des données saisies.....	62
Nettoyage des données.....	62
Recodage général.....	62
Examen des données d'échantillonnage.....	62
Dernier examen des données et préparation de la base de données.....	63
Chapitre 9. Analyse des données sur le rendement des élèves.....	64
Analyse préliminaire.....	64
Filtrage des données.....	64
Recodage des items.....	65
Données manquantes.....	66
Items non utilisés dans l'évaluation.....	66
Items sans objet.....	67
Items non atteints.....	67
Items omis.....	67
Réponse non valable.....	68
Analyse des items.....	68
Analyse de la théorie classique des items.....	68
Difficulté des items.....	69
Statistiques spécifiques pour les items à RCS.....	69
Examen pour mettre en évidence les données manquantes.....	69
Fiabilité de l'évaluation du PPCE 2016.....	70
Items problématiques.....	70
Analyse de la TRI.....	70
Évaluation de la dimensionnalité du PPCE 2016.....	71
Calibrage des items et évaluation du caractère convenable des modèles de la TRI.....	71
Fonctionnement différentiel des items.....	72
Fonctionnement du test.....	72
Liens et équivalences avec les évaluations précédentes pour les domaines secondaires.....	72
Production des scores pour le rendement et scores à l'échelle.....	73
Estimations des erreurs-types.....	74
Présentation des résultats de l'évaluation du PPCE 2016.....	74
Chapitre 10. Analyse des données du questionnaire.....	75
Analyse préliminaire.....	75
Filtrage des données.....	75

Recodage des items	76
Données manquantes	76
Statistiques descriptives	77
Analyse factorielle.....	77
Analyses des items et des indices.....	77
Analyse de comparaison des groupes	78
Analyse corrélacionnelle	78
Chapitre 11. Ensembles de données du PCE	79
Description des ensembles de données	79
Ensemble de données des élèves	79
Ensemble de données du personnel enseignant.....	79
Ensemble de données des écoles	80
Ensemble de données fusionné – élèves/personnel enseignant/écoles	80
Consultation de l'ensemble de données pour les recherches.....	80
Conditions et modalités	81
Information pour les chercheuses et chercheurs	82
Bibliographie	83

Liste des tableaux

Tableau 2.1 – Calendrier des évaluations du PPCE.....	12
Tableau 2.2 – Nombre de blocs, de scénarios et d’items par domaine et par livret	14
Tableau 2.3 – Répartition des items selon le type et le domaine d’évaluation	14
Tableau 2.4 – Lignes directrices à l’intention du Groupe consultatif sur l’équité des tests	19
Tableau 4.1 – Couverture de la population cible du PPCE	26
Tableau 4.2 – Répartition de la taille des échantillons du PPCE 2016.....	28
Tableau 4.3 – Estimation de la taille d’un échantillon.....	29
Tableau 4.4 – Résumé des critères utilisés pour échantillonner les élèves, selon les types de strates.....	32
Tableau 6.1 – Résultats des examens de fidélité pour le groupe de correction en lecture n° 1..	47
Tableau 6.2 – Résultats des examens de fidélité pour le groupe de correction en lecture n° 2..	47
Tableau 6.3 – Résultats des examens de fidélité pour le groupe de correction en mathématiques	48
Tableau 6.4 – Résultats des examens de fidélité pour le groupe de correction en sciences.....	48
Tableau 6.5 – Niveau global d’accord entre membres du groupe de correction pour la double correction	48
Tableau 7.1 – Composition des livrets d’évaluation du PPCE pour les items de lecture	56
Tableau 7.2 – Répartition des élèves par niveau de rendement en lecture	59

Chapitre 1. Vue d'ensemble du Programme pancanadien d'évaluation

Contexte

Depuis une vingtaine d'années, les ministères de l'Éducation participent à diverses évaluations pour mesurer les compétences des élèves en raisonnement, résolution de problèmes et communication afin de mieux les préparer pour l'avenir. Sur la scène internationale, par l'entremise du Conseil des ministres de l'Éducation (Canada) [CMEC], ils ont participé au Programme international pour le suivi des acquis des élèves (PISA) de 2000, 2003, 2006, 2009, 2012, 2015 et 2018 (avec plus de 80 pays et économies en 2018), au Programme international de recherche en lecture scolaire (PIRLS) de 2011 et 2016 (avec plus de 60 pays), à l'évaluation Tendances de l'enquête internationale sur les mathématiques et les sciences (TEIMS) de 2015 et 2019 (avec environ 65 pays) et à l'Enquête internationale sur la maîtrise de l'ordinateur et de l'information (EIMOI) de 2013 (avec environ 20 pays). En outre, la plupart des provinces/territoires procèdent à leurs propres évaluations du rendement de leurs élèves à différentes étapes de leur scolarité. Dans l'optique d'examiner le contexte pédagogique, certaines provinces ont participé, par l'entremise du CMEC, à l'évaluation Teacher Education and Development Study – Mathematics – TEDS-M (étude sur l'éducation et le perfectionnement du personnel enseignant les mathématiques) en 2008 et à l'Enquête internationale sur les enseignants, l'enseignement et l'apprentissage (TALIS) de 2013. L'évaluation du Programme pour l'évaluation internationale des compétences des adultes (PEICA) s'est déroulée en 2012 afin d'étudier de façon générale le niveau des adultes en littératie, en numératie et en résolution de problèmes dans 25 pays, dont le Canada. La population canadienne souhaite depuis longtemps savoir dans quelle mesure les systèmes éducatifs répondent aux besoins des élèves et de la société.

Évaluation pancanadienne

Afin d'étudier et de rendre compte du rendement des élèves dans un contexte canadien, le CMEC a créé, en 1989, le Programme d'indicateurs du rendement scolaire (PIRS) pour évaluer le rendement des élèves de 13 et 16 ans au Canada. Le PIRS était un programme cyclique d'évaluations pancanadiennes des compétences des élèves en lecture et écriture, en mathématiques et en sciences, effectuées entre 1993 et 2004. En 2003, les ministres de l'Éducation des provinces et des territoires se sont entendus, par l'entremise du CMEC, pour mettre sur pied le PPCE en vue de remplacer le PIRS. Le domaine principal de chaque évaluation du PPCE est l'une de ces disciplines d'apprentissage, mais chaque évaluation comprend les deux autres disciplines en tant que composantes secondaires.

Les programmes d'études diffèrent d'une région à l'autre du pays, de sorte qu'il est difficile d'en comparer les résultats. Toutefois, les jeunes Canadiennes et Canadiens des différentes provinces et des différents territoires acquièrent des habiletés similaires en lecture, en mathématiques et en sciences. Le PPCE a été conçu pour déterminer si les élèves de tout le Canada atteignent un niveau de rendement semblable dans ces matières de base à un âge à peu près identique et sert de complément aux évaluations existantes qu'effectuent les

provinces et les territoires en leur apportant des données pancanadiennes comparatives sur les niveaux atteints par les élèves de 8^e année/2^e secondaire de tout le pays.

Le PPCE est une évaluation à l'échelle du système et est destiné principalement aux ministères provinciaux responsables de l'éducation pour qu'ils examinent leurs systèmes d'éducation. Les renseignements rassemblés grâce à chaque évaluation fournissent aux ministres de l'Éducation la base dont ils ont besoin pour examiner les programmes d'études et les autres aspects des systèmes scolaires. Les données du PPCE sont présentées pour chaque province et chaque territoire, selon la langue du système scolaire et selon le sexe. Les évaluations pancanadiennes et internationales de grande envergure visent à fournir de l'information fiable sur le rendement scolaire et à aider à comprendre les facteurs contextuels qui influent sur ce rendement. Elles fournissent aux responsables de l'élaboration des politiques, aux administratrices et administrateurs, au personnel enseignant et au milieu de la recherche de l'information précieuse sur le fonctionnement des systèmes d'éducation ainsi que sur les façons possibles de les améliorer. Il importe toutefois de signaler que l'évaluation n'est pas conçue pour produire des résultats valides à l'échelon des élèves, des écoles ou des conseils/commissions/districts scolaires.

En 2007, ce sont, pour commencer, les élèves âgés de 13 ans qui ont participé au PPCE. Depuis 2010, cependant, l'évaluation du PPCE concerne les élèves de 8^e année/2^e secondaire et dans la mesure du possible, des classes entières sont sélectionnées afin de minimiser les dérangements pour les classes et les écoles.

L'évaluation du PPCE *ne porte pas* sur le rendement individuel de l'élève. Elle ne fait pas non plus intervenir de comparaisons entre élèves, entre écoles ou entre conseils/commissions/districts scolaires. Les résultats du PPCE ne sont pas mis à la disposition du personnel enseignant, des conseils/commissions/districts scolaires, des régions ou des ministères de l'Éducation pour qu'ils puissent s'en servir pour évaluer les résultats obtenus par les écoles fréquentées par les élèves en question.

Participation

En 2007 et en 2010, 10 provinces et un territoire (le Yukon) ont participé aux deux premières éditions du PPCE. En 2013 et en 2016, 10 provinces ont participé au PPCE. Les Territoires du Nord-Ouest ont, par le passé, participé au PIRS.

Durée de l'évaluation

Les élèves se sont vu accorder 90 minutes pour répondre aux items de l'évaluation du PPCE. Ils ont eu droit à 30 minutes supplémentaires pour terminer le test, si nécessaire. Il était possible d'accorder du temps supplémentaire aux élèves pour lesquels ce type d'adaptation était prévu dans le programme scolaire ordinaire. Après avoir répondu aux items du livret d'évaluation, les élèves ont eu 30 minutes pour répondre au questionnaire qui leur était destiné. Les élèves étaient autorisés à utiliser les ressources auxquelles ils avaient normalement accès en cours de langue, de mathématiques et de sciences. Le Questionnaire du personnel enseignant et le

Questionnaire de l'école ont également servi à se faire une idée plus globale des systèmes éducatifs au Canada.

Le PPCE dans les deux langues officielles

Les résultats du Canada francophone sont ceux des élèves qui fréquentent le système d'éducation de langue française de leur province respective. De la même manière, les résultats du Canada anglophone sont ceux des élèves qui fréquentent le système d'éducation de langue anglaise de leur province respective. La majorité des élèves des programmes d'immersion en français ont passé l'évaluation en lecture en anglais plutôt que dans leur langue seconde. La liste des termes scientifiques et mathématiques courants en anglais et en français a été fournie comme ressource à ces élèves.

Chapitre 2. Conception et élaboration de l'évaluation

Afin d'éviter tout biais d'ordre linguistique, l'instrument d'évaluation du PPCE a été conçu parallèlement en français et en anglais par des spécialistes francophones et anglophones de l'éducation. Tous les items des trois disciplines ont été rédigés dans les deux langues et tous les élèves concernés par la mise à l'essai ou l'étude principale du PPCE ont répondu aux mêmes questions, indépendamment de la langue. Les échantillons retenus dans le cadre du PPCE ont été choisis afin qu'ils soient représentatifs à la fois du groupe de langue officielle majoritaire et du groupe de langue officielle minoritaire dans les huit provinces comptant un nombre suffisant d'individus pour qu'il soit possible de faire des comparaisons statistiques valables. En raison de la petite taille de l'échantillon, les résultats des élèves inscrits dans les écoles francophones de l'Île-du-Prince-Édouard ne figurent pas dans les résultats du PPCE; en revanche, ils ont été inclus dans le calcul du résultat global de cette province. L'échantillon d'élèves francophones pour la Saskatchewan était certes lui aussi de très petite taille, mais il représentait la majorité de l'effectif d'élèves francophones de 8^e année dans la province. Le ministère de l'Éducation de la Saskatchewan a approuvé l'inclusion des données sur cette population dans les rapports sur les résultats.

Les élèves francophones de Terre-Neuve-et-Labrador n'ont pas participé au PPCE 2016.

Cycle d'évaluation du PPCE

Les évaluations du PPCE sont menées tous les trois ans auprès des élèves de 8^e année/2^e secondaire. Chaque cycle d'évaluation collecte des données sur le rendement à l'aide de tests cognitifs qui ciblent principalement l'un des trois domaines d'apprentissage – la lecture, les mathématiques et les sciences –, mais qui s'intéressent également de façon secondaire aux deux autres. En outre, le PPCE collecte un vaste éventail de renseignements contextuels (par exemple, sur le contexte démographique, les facteurs socioéconomiques, l'enseignement donné à l'école et les conditions d'apprentissage) pour éclairer l'interprétation du rendement des élèves.

Chaque évaluation du PPCE comporte des questions sur les trois domaines, mais le domaine principal varie d'un cycle à l'autre, comme le montre le tableau 2.1 ci-après. La réalisation de ces évaluations à intervalles réguliers produit des données d'actualité pouvant être comparées entre les provinces et les territoires, et au fil du temps. En 2016, la quatrième évaluation était axée sur la lecture, comme cela avait été le cas pour la première évaluation en 2007 – les mathématiques et les sciences constituant les domaines secondaires.

Tableau Error! No text of specified style in document..1 – Calendrier des évaluations du PPCE

Domaine	Dates réelles et proposées pour les évaluations du PPCE					
	Cycle 1			Cycle 2		
	Printemps 2007	Printemps 2010	Printemps 2013	Printemps 2016	Printemps 2019	Printemps 2022
Principal	Lecture	Mathématiques	Sciences	Lecture	Mathématiques	Sciences
Secondaire	Mathématiques	Sciences	Lecture	Mathématiques	Sciences	Lecture
Secondaire	Sciences	Lecture	Mathématiques	Sciences	Lecture	Mathématiques

Présentation des résultats du PPCE au fil du temps

L'un des points forts du PPCE est qu'il mesure l'évolution du rendement des élèves au fil du temps. Les échelles de rendement du PPCE offrent un point de référence commun qui permet aux provinces et aux territoires de comparer, d'une évaluation à l'autre, les progrès des élèves de 8^e année/2^e secondaire dans les trois matières de base. Les items utilisés lors des années de référence, appelés « items d'ancrage », serviront de base à l'établissement de liens entre les résultats des diverses évaluations. Cette base permettra ainsi aux provinces et aux territoires de comparer les données de 2007, 2010, 2013 et 2016, et de déterminer comment le rendement évolue au fil du temps.

En 2010, la définition de la population cible est passée d'une définition fondée sur l'âge (13 ans) à une définition fondée sur l'année d'études (8^e année/2^e secondaire). Comme les résultats ont été échelonnés séparément dans les deux évaluations, avec une moyenne de 500 et un écart-type de 100, il s'est avéré nécessaire de reporter les résultats échelonnés de l'évaluation de 2007 sur l'échelle de l'évaluation de 2010. Ce rééchelonnage a entraîné des variations des moyennes de 2007 indiquées pour la lecture entre les deux rapports, publiés respectivement en 2007 et en 2010.

Mise à jour du cadre d'évaluation

La mise à jour du cadre d'évaluation du PPCE 2016 a commencé par l'examen et la modification des cadres d'évaluation qui précisent le contenu à évaluer. Bien que les programmes scolaires diffèrent d'un bout à l'autre du pays, le PPCE se fonde sur les domaines communs à tous les programmes d'études de 8^e année/2^e secondaire. Cet accent mis sur les domaines communs permet donc d'établir des comparaisons entre les élèves des diverses provinces à une étape comparable de leur scolarité. Le *Cadre d'évaluation du PPCE 2016* (CMEC, 2016)³ présente la théorie, les principes de conception et les descripteurs de rendement utilisés dans l'élaboration des items des trois domaines pour le deuxième cycle du PPCE (2016-2022). Le chapitre 2 décrit

³ https://www.cmec.ca/553/Documents_d_information.html

le domaine principal, soit la lecture, tandis que les chapitres 3 et 4 décrivent les domaines secondaires, soit les mathématiques et les sciences, respectivement.

Pour l'année 2016, les cadres d'évaluation de la lecture et des mathématiques ont été mis à jour pour mieux refléter les programmes d'études et les normes à l'échelle du Canada. Le cadre d'évaluation des sciences avait été mis à jour pour le PPCE 2013, alors que les sciences avaient été pour la première fois le domaine principal; par conséquent, le contenu du cadre est resté inchangé depuis ce cycle. Les spécialistes provinciaux ont discuté des mises à jour et y ont apporté des révisions.

Un nouveau domaine de contenu a été ajouté au cadre d'évaluation de la lecture de 2016. Le monde de l'information numérique exige que les lectrices et lecteurs se servent de leur habileté et fassent un effort pour acquérir des connaissances cohérentes à partir de nombreux éléments de médias, y compris le texte et les images. Cette capacité de construire des connaissances cohérentes représente un nouveau défi pour la littératie, allant au-delà de tout ce qui s'est présenté antérieurement, et c'est ce que nous appelons la « translittératie ». Un élément contextuel du PPCE 2016 comportait l'introduction d'une tâche complexe qui exigeait de multiples représentations de l'information (c'est-à-dire sous forme de texte, de diagramme, de graphique ou de carte) et l'évaluation mettait l'accent sur la production par l'élève d'une explication cohérente de l'événement, du phénomène ou du problème central.

Conception de l'évaluation

L'évaluation du PPCE est un test papier-crayon qui couvre trois domaines d'évaluation : la lecture, les mathématiques et les sciences. En 2016, le domaine principal était la lecture, tandis que les mathématiques et les sciences étaient les domaines secondaires. Tout comme dans le PISA, le domaine principal change à chaque édition de l'évaluation; cela signifie que, en 2019, la lecture sera un domaine secondaire et ce sont les mathématiques qui seront le domaine principal du PPCE.

Pour l'évaluation du PPCE, les unités d'évaluation en lecture ont été regroupées en huit groupes ou « blocs » répartis sur quatre livrets, de sorte que chaque livret contenait généralement deux blocs d'items en lecture, un bloc d'items en mathématiques et un bloc d'items en sciences. Un livret contenait un bloc de questions multidisciplinaires qui servaient à mesurer les connaissances et les compétences des élèves en lecture, en mathématiques et en sciences dans un seul et même contexte. Les quatre livrets ont été distribués aléatoirement et de façon égale aux élèves d'une même classe. Par conséquent, chaque élève a répondu à deux des huit blocs d'items d'évaluation en lecture; toutefois, les huit blocs ont été effectués par l'ensemble des élèves d'une classe. De plus, des paires de livrets comportaient des ensembles ou des unités d'items communs, permettant de faire des mesures comparatives du rendement des élèves d'un livret à l'autre. Tous les livrets d'évaluation contenaient un questionnaire de l'élève à la fin.

Le tableau 2.2 montre la répartition des blocs, des contextes (ou scénarios ou passages) et des items pour la lecture, les mathématiques et les sciences dans l'ensemble des quatre livrets, tandis que le tableau 2.3 montre la répartition des types d'items pour les trois domaines.

Tableau 2.2 – Nombre de blocs, de scénarios et d’items par domaine et par livret

	Lecture			Mathématiques			Sciences		
	Blocs	Passages	Items	Blocs	Scénarios	Items	Blocs	Scénarios	Items
Livret 1	2	4	25	1	4	8	1	3	9
Livret 2	2	5	25	1	4	9	1	3	9
Livret 3	2	6	26	1	4	9	1	3	7
Livret 4	2	5*	17	1	4*	12	1	4*	12

*inclut un scénario multidisciplinaire

Tableau 2.3 – Répartition des items selon le type et le domaine d’évaluation

Domaine	Livret 1		Livret 2		Livret 3		Livret 4	
	RC	RCS	RC	RCS	RC	RCS	RC	RCS
Lecture	19	6	20	5	18	8	15	3
Mathématiques	4	4	5	4	5	4	7	5
Sciences	7	2	5	4	4	3	9	3
Total	30	12	30	13	27	15	31	11

Groupes de travail

Le domaine principal du PPCE 2016 était la lecture. Ainsi, de nouveaux items cognitifs ont été élaborés pour ce domaine seulement, et les questionnaires ont été révisés pour mettre l’accent sur l’enseignement et l’apprentissage de la lecture. Les groupes de travail se composaient de spécialistes des programmes d’études des domaines langagiers, ainsi que de l’enseignement, de l’apprentissage et de l’évaluation. Ils provenaient de diverses provinces et près de la moitié d’entre eux étaient bilingues. Ces spécialistes ont travaillé de façon approfondie sur le PPCE et ont participé à diverses étapes du projet, dont l’élaboration du cadre d’évaluation, la préparation des premières versions des items, la validation et la révision des items et la comparaison entre les items en français et les items en anglais. Certains ont également participé aux séances de correction des items lors de la mise à l’essai de l’évaluation et lors de l’étude principale.

Mise à jour du cadre d’évaluation en lecture

M. Carl Bereiter et M^{me} Marlene Scardamalia, de la Knowledge Building Concepts Foundation, Incorporated (fondation pour les concepts de développement des connaissances, incorporée), ont été chargés de mettre à jour le cadre d’évaluation en lecture. Ce cadre avait été élaboré au début du déploiement du PPCE en 2007 et se basait sur la recherche actuelle liée à l’enseignement de la littératie. Le cadre révisé reflète la pensée actuelle dans le domaine de la recherche sur la littératie, y compris l’évaluation en lecture.

Le groupe de travail chargé de l’examen du cadre d’évaluation en lecture s’est penché sur ledit cadre et y a apporté des révisions supplémentaires pour refléter les changements dans les programmes d’études des domaines langagiers en vigueur dans les provinces de l’ensemble du Canada. Le groupe de travail, qui s’est réuni à Toronto en juin 2014, comprenait des spécialistes

des programmes d'études dans les domaines langagiers et de l'évaluation, en provenance de l'Alberta, du Québec, du Nouveau-Brunswick et de l'Île-du-Prince-Édouard.

Élaboration des items

Les documents devant servir à guider toutes les étapes du processus d'élaboration des items ont été préparés pour la réunion des responsables de l'élaboration des tests, qui a eu lieu à Toronto en août 2014.

Les provinces et les territoires ont été invités à nommer des responsables de l'élaboration des items et ce groupe de travail se composait de représentantes et représentants de la Colombie-Britannique, de l'Alberta, de la Saskatchewan, du Manitoba, de l'Ontario, du Québec, du Nouveau-Brunswick, de la Nouvelle-Écosse, de l'Île-du-Prince-Édouard et de Terre-Neuve-et-Labrador.

Le processus d'initiation comprenait une vue d'ensemble du cadre d'évaluation en lecture, du processus d'élaboration des items et des échéances, une description précise des exigences pour les items et des remarques insistant sur l'importance de la conformité au cadre d'évaluation. La séance a commencé par une discussion en grand groupe en vue de mettre en évidence des sujets susceptibles d'intéresser les élèves de 8^e année/2^e secondaire et de correspondre à l'ensemble des programmes d'études au Canada pour le groupe d'âge concerné. L'élaboration des items s'est faite par petits groupes et s'est déroulée simultanément en anglais et en français. Les séances ont fait intervenir un processus itératif : de petits groupes travaillaient à l'élaboration d'une unité contenant une série de questions autour d'un stimulus correspondant au cadre d'évaluation. Pour avoir une unité complète, il fallait le texte de stimulus, une série de quatre à sept items combinant des items à réponse choisie et des items à réponse construite et un guide pour le codage des réponses à chaque question. Chaque guide de codage consistait en une liste de catégories de réponses (crédit complet, crédit partiel, pas de crédit), chacune ayant son propre code, des descriptions des types de réponses auxquelles chaque code serait attribué et des exemples de réponses dans chaque catégorie. Les unités ont été présentées au grand groupe pour une discussion sur la qualité des items, la pertinence par rapport à l'âge des élèves, la prise en compte des différences culturelles et des différences entre les sexes, la couverture du programme d'études et la correspondance avec le cadre d'évaluation. Après la discussion, les petits groupes ont révisé leurs items en tenant compte des suggestions et des recommandations du grand groupe. À la fin du premier cycle d'élaboration des unités, le grand groupe s'est à nouveau rassemblé pour choisir le sujet suivant pour le stimulus, afin de s'assurer que le cadre d'évaluation en lecture était bien couvert.

À la fin de cette séance d'élaboration des items, le groupe de travail a réexaminé et révisé le cadre d'évaluation en lecture pour qu'il corresponde bien aux sujets et aux types de questions susceptibles de représenter les points communs entre les différents programmes d'études de 8^e année/2^e secondaire au Canada. De petits groupes ont ensuite réexaminé les items de lecture des évaluations antérieures pour s'assurer qu'ils représentaient bien le cadre d'évaluation et avaient bien la bonne classification.

Traduction et examen des items

Les items ont été élaborés à la fois en français et en anglais et ont fait l'objet d'un travail de traduction et de révision au CMEC.

Un groupe de travail chargé d'examiner les items s'est réuni à Toronto en janvier 2015. Ce groupe de travail était composé de représentantes et représentants de l'Ontario, du Nouveau-Brunswick et de l'Île-du-Prince-Édouard. Le but de la rencontre était d'examiner les items de lecture pour vérifier le contenu, le vocabulaire, la traduction, la conformité aux programmes d'études et l'absence de tout biais et aussi pour vérifier la classification des items pour les sous-domaines décrits dans le cadre d'évaluation. Le groupe de travail a fait, pour chaque item, l'une des trois recommandations suivantes : a) garder l'item tel quel; b) enlever l'item de la banque d'items pour la mise à l'essai; ou c) garder l'item en y apportant des changements mineurs recommandés par les membres du groupe de travail. À l'issue du travail de ce groupe de travail, les unités achevées contenaient le texte de stimulus et entre deux et six items. Les unités pouvaient contenir un seul type d'item (à réponse choisie ou à réponse construite) ou un mélange des deux types d'items.

À la fin de la rencontre de ce groupe de travail, le cadre d'évaluation a une nouvelle fois été réexaminé et révisé afin de mieux tenir compte des éléments communs des différents documents décrivant les programmes d'études des provinces. Dans certains cas, les items ont été éliminés en raison de biais relatifs au sexe ou à la culture des individus ou parce qu'ils posaient problème après traduction. Les unités restantes ont été révisées et vérifiées dans les deux langues.

Vérification des items d'évaluation et des guides de codage

Avant d'inclure les items dans une évaluation, que ce soit pour la mise à l'essai ou pour l'étude principale du PPCE, il était important que ces items soient examinés selon divers points de vue par des groupes de spécialistes pour s'assurer que les items étaient raisonnables et fournissaient une évaluation exacte des compétences des élèves de 8^e année/2^e secondaire partout au Canada. Le processus de validation, mis en œuvre par les groupes de spécialistes, comprenait les étapes suivantes :

- traduction et comparaison des items en français et en anglais;
- assurance de l'équivalence des items dans les deux langues par rapport au niveau de difficulté;
- vérification de la classification des items de lecture conformément au cadre d'évaluation en lecture;
- vérification des guides de correction pour les items à réponse construite;
- révisions linguistiques et stylistiques, révisions techniques et révisions psychométriques.

Traduction et comparaison des items en français et en anglais

Les unités d'items élaborées dans les deux langues officielles par le groupe de travail chargé de l'élaboration des items ont fait l'objet d'une traduction croisée par les traductrices et traducteurs du CMEC.

Dans une évaluation à grande échelle comme le PPCE, il est vital de s'assurer qu'il y a une bonne concordance entre les diverses versions du test sur le plan linguistique, afin d'éviter de donner un avantage à un groupe particulier par rapport à un autre. Même s'il risque toujours d'y avoir des différences entre items dans une évaluation, il est nécessaire de s'assurer que les items de la version française sont aussi équivalents que possible aux items de la version anglaise. En outre, tout texte suppose que les élèves ont un certain niveau en littérature pour la lecture. Dans l'évaluation du PPCE, les contextes et les passages ont été choisis afin qu'ils soient accessibles à la grande majorité des élèves de 8^e année/2^e secondaire. Des groupes de travail bilingues composés d'éducatrices et éducateurs expérimentés ont examiné et validé les items à chaque stade de leur élaboration, afin de s'assurer que le vocabulaire correspondait bien au niveau de compréhension attendu de la part des élèves du Canada à ce niveau.

Révisions linguistiques et typographiques

Les révisions linguistiques et typographiques sont une étape importante de la révision des items. La correction linguistique devait porter sur des aspects comme la grammaire, la syntaxe, l'orthographe et la ponctuation pour chaque item, scénario ou graphique dans chaque livret. Il a ensuite fallu effectuer une révision typographique sur des aspects comme les espaces, les polices de caractères, le nombre de lignes, la mise en page et les introductions des énoncés. Les réviseuses et réviseurs ont dû vérifier que la taille des caractères était la même dans tous les items, que l'interligne était constant dans tous les livrets, que la mise en page était cohérente, que chaque item commençait par un énoncé suivi d'une question, que le nombre de lignes prévu pour la réponse de l'élève correspondait bien à la longueur de la réponse attendue, et que les références étaient exactes (autrement dit, que, lorsqu'un item faisait référence à un graphique d'une autre page, le numéro de page était correct).

Révisions techniques

Les révisions techniques servent à vérifier et à valider les réponses correctes, les calculs, les données, etc. Les quatre versions du test contenaient entre autres des items à réponse choisie, chacun avec quatre réponses possibles. Il fallait s'assurer et vérifier qu'il y avait bien une seule réponse correcte et que les trois autres choix étaient des leurres raisonnables sur le plan logique. En sciences et en mathématiques, les items pouvaient exiger que les élèves fassent des calculs pour obtenir la réponse correcte. Il fallait donc refaire les calculs pour vérifier que la réponse définitive était bien une des réponses au choix. Pour les items à réponse construite, il n'y avait pas de réponse à vérifier, mais il fallait malgré tout valider à nouveau les items (et les exemples de réponses) afin de s'assurer qu'on avait bien attribué les descripteurs corrects et vérifié leur exactitude, soit en revenant au texte soit en faisant les calculs.

Plusieurs questions ou scénarios de mathématiques et de sciences comprenaient des tableaux, des diagrammes et des graphiques avec des données. Les réviseuses et réviseurs devaient donc vérifier et garantir l'exactitude des informations. Il était aussi possible que les élèves aient à se référer à un tableau ou à un graphique pour trouver la réponse correcte. L'item indiquait aux élèves à quelle page le tableau ou le graphique en question se trouvait. Les réviseuses et réviseurs devaient donc s'assurer que le numéro de la page en question était correct.

Plusieurs questions de lecture avaient des numéros de ligne ou de paragraphe. Les réviseuses et réviseurs ont vérifié que le système de numérotation était cohérent dans les différentes versions du test. Dans le cas des items d'ancrage, ils ont également vérifié que les items étaient bien identiques dans les livrets des différentes évaluations du PPCE.

Lors de la révision des items, il était important de vérifier que tous les volets du texte ou de l'item étaient bien présents, de façon à ce que les élèves puissent bel et bien répondre à la question. Si certains volets manquaient dans l'item, par exemple, les élèves seraient dans l'incapacité de fournir une réponse correcte à la question et il serait alors nécessaire d'exclure l'item en question des analyses. Il aurait été malheureux d'avoir à enlever un item du test, en particulier s'il était utile pour mesurer les compétences des élèves.

Révisions psychométriques

Les spécialistes en sciences, en lecture et en mathématiques ont également eu à faire une révision psychométrique des items. Pour les items à réponse choisie, l'un des facteurs à vérifier était l'ordre des réponses possibles. En lecture, les réponses possibles commençaient par la plus courte et se terminaient par la plus longue, c'est-à-dire de la phrase la plus courte à la phrase la plus longue. Lorsque les réponses possibles étaient des nombres, les leurres pouvaient être présentés par ordre croissant, du plus petit au plus grand. Cette approche permettait de présenter la réponse correcte à un endroit aléatoire dans la liste. Il fallait que chaque réponse possible soit approximativement de la même longueur. Si l'une des réponses était plus détaillée, les élèves auraient davantage tendance à la choisir et à donner la réponse correcte. Il était également important de vérifier l'exactitude des réponses correctes, afin de s'assurer qu'il n'y avait pas d'autre réponse susceptible d'être également correcte et d'éviter ainsi toute ambiguïté.

Les spécialistes des domaines langagiers ont élaboré un guide de codage avec des descripteurs pour les nouveaux items à réponse construite. Les guides de codage pour les items d'ancrage sont restés inchangés par rapport à l'évaluation précédente dans laquelle ces matières étaient le domaine principal (soit l'évaluation de 2007 pour la lecture, l'évaluation de 2010 pour les mathématiques et l'évaluation de 2013 pour les sciences), afin de garantir la cohérence des corrections des items et la possibilité d'analyser les changements dans les résultats au fil du temps. Divers codes ont été attribués aux réponses des élèves. En mathématiques et en sciences, les codes pouvaient être de 0 ou 1 ou de 0, 1 ou 2. En lecture, les codes allaient de 0 à 3. Chaque code comprenait une description complète et un ou plusieurs exemples tirés des réponses des élèves. Les spécialistes ont donc dû passer en revue tous les critères de correction et s'assurer que les codes définis étaient clairs et précis. Cette étape était très importante, parce que, lors de la séance de correction des items dans les trois matières, les correctrices et

correcteurs avaient reçu une formation sur chaque item à coder. Il fallait qu'ils soient en mesure de faire clairement la distinction entre les différents codes, afin de pouvoir attribuer à chaque réponse d'élève le code le plus approprié.

Les spécialistes ont également dû passer en revue les tableaux de spécifications, qui présentaient le plan principal de l'évaluation, et valider les types d'items. Il fallait, par exemple, que l'évaluation comprenne un mélange équilibré d'items à réponse construite et d'items à réponse choisie, afin de bien utiliser le temps consacré par les élèves à l'évaluation, tout en recueillant les réactions critiques et personnelles dans un contexte ouvert.

Groupe consultatif sur l'équité des tests

À la 106^e réunion du CCSME en juin 2015, les sous-ministres ont approuvé la création d'un groupe consultatif visant à fournir de la rétroaction sur la pertinence des documents d'évaluation utilisés dans les évaluations pancanadiennes et internationales effectuées avec le concours du CMEC. L'objectif principal du travail de ce groupe est de veiller à ce que les items des questionnaires et les passages ou scénarios utilisés pour les items d'évaluation ne présentent pas de biais et soient le plus valides et équitables possible pour un vaste éventail d'élèves.

Le principe directeur du groupe était de veiller à ce qu'aucun élément d'une évaluation ne fasse en sorte qu'un élève soit dérangé ou distrait à un point tel qu'il ne soit pas en mesure de communiquer son attitude ou sa compréhension du sujet faisant l'objet de l'évaluation. Les contextes et les items des questionnaires d'évaluation devaient tenir compte des valeurs de l'éducation publique inclusive de façon à respecter la diversité au chapitre de la race, de la couleur, de la religion, de l'origine nationale, de l'ascendance, du lieu d'origine, de l'âge, de l'incapacité, de l'état civil, de l'orientation sexuelle ou de l'identité sexuelle réelle ou perçue, du sexe, de la condition sociale, des convictions ou activités politiques.

Il a été demandé au groupe consultatif de considérer neuf éléments dans son examen des contextes et des items des questionnaires, comme le montre le tableau 2.4.

Tableau 2.4 – Lignes directrices à l'intention du Groupe consultatif sur l'équité des tests

LIGNES DIRECTRICES À L'INTENTION DU GROUPE CONSULTATIF SUR L'ÉQUITÉ DES TESTS	
1. Universalité	<ul style="list-style-type: none">• Les items sont appropriés pour tous les élèves, les enseignantes et enseignants, les directrices et directeurs ainsi que les parents participants (y compris les sous-populations).• Les items n'utilisent pas de mots, de tournures et d'expressions idiomatiques inconnus.• Les items n'utilisent pas un langage ou des expressions pouvant être choquants pour un groupe particulier.
2. Sexe	<ul style="list-style-type: none">• Les items ne comprennent pas de thèmes, de sujets ou un ton organisationnel favorisant un sexe.• Les items ne brossent pas le portrait des rôles des sexes de façon stéréotypée ou

péjorative.

3. Géographie

- Les items ne donnent pas d'avantages ou de désavantages aux élèves de certaines régions du Canada.

4. Situation socio-économique

- Les items ne mettent pas l'accent sur la richesse, la pauvreté et le crime de façon inappropriée.
- Les items ne laissent pas entendre que la richesse ou la pauvreté est attribuée à un groupe particulier.
- Les items ne laissent pas entendre que l'appartenance à un groupe socio-économique particulier est plus avantageuse que l'appartenance à un autre groupe.

5. Situation sociale et environnementale

- Les items ne font pas référence à des questions touchant la violence, les jeux de hasard, les dépendances, les comportements criminels ou l'itinérance.
- Les items évitent les sujets comme les conflits familiaux.
- Les items évitent de mettre l'accent sur la guerre et la vie des soldats (d'hier ou d'aujourd'hui).

6. Race, ethnie et culture

- Les items ne représentent pas de groupes ethniques de façon péjorative ou stéréotypée.
- Les items ne comprennent pas de sujets qui laissent entendre qu'une culture est préférée ou supérieure à une autre.
- Les items ne comprennent pas de références iconographiques culturelles à des célébrations, à des fêtes nationales (autres que canadiennes) ou à des concours/prix.

7. Religion

- Les items évitent des sujets qui exigent que les élèves prennent une position qui pourrait aller à l'encontre de leurs croyances ou enseignements religieux.
- Les items évitent les références religieuses.

8. Incapacité

- Les items présentent de façon naturelle et authentique des personnes ayant des incapacités.
- Les items ne mettent pas l'accent sur l'incapacité ou sur les suppositions au sujet des défis perçus que doit relever une personne.

9. Âge

- Les items ne présentent pas un groupe d'âge de façon péjorative ou stéréotypée.
- Les items ne présentent pas un groupe d'âge comme ayant un avantage sur un autre.

Les membres du Groupe consultatif sur l'équité des tests ont été invités à examiner les documents d'évaluation pour veiller à ce qu'il n'y ait pas de biais lié à leur domaine d'expertise précis et à fournir des directives à l'équipe responsable de l'élaboration afin de raffiner les items. Ils ont formulé des recommandations concernant les items qui devraient être approuvés sans changement, ceux qui devraient être approuvés avec des changements ou ceux qui devraient être rejetés. Ces recommandations ont servi à préparer les ébauches finales des documents d'évaluation.

Approbation des items par les provinces

Avant d'inclure les items dans la mise à l'essai et l'étude principale, il fallait obtenir l'approbation par les provinces des items choisis. Le CMEC a produit trois livrets pour la mise à l'essai et quatre livrets pour l'étude principale, en français et en anglais, et les a envoyés aux provinces pour qu'elles les examinent. Le CMEC a obtenu l'approbation de chaque province pour inclure les scénarios ou passages et les items dans la mise à l'essai en 2015 et l'étude principale en 2016.

Chapitre 3. Élaboration des questionnaires contextuels

Les élèves qui participent au PPCE ainsi que le personnel enseignant et la direction de leur école remplissent des questionnaires conçus pour fournir aux provinces et aux territoires de l'information contextuelle facilitant l'interprétation des résultats en matière de rendement. Le milieu de la recherche, les responsables de l'élaboration des politiques et les praticiennes et praticiens peuvent eux aussi se servir de l'information recueillie grâce à ces questionnaires pour déterminer quels facteurs influent sur les résultats d'apprentissage. Le contenu des questionnaires contextuels varie selon laquelle des trois matières est évaluée comme domaine principal du PPCE.

Les questions contextuelles qui accompagnent le PPCE 2016 portent notamment sur des facteurs qui, selon des études antérieures, sont en corrélation avec le rendement en lecture. Les facteurs suivants n'en sont que quelques exemples :

- niveau de scolarité des parents;
- langue parlée à la maison;
- nombre de livres à la maison.

Les questionnaires contextuels portent également sur les conditions d'enseignement et d'apprentissage, y compris les attentes des enseignantes et enseignants au chapitre des devoirs, leurs domaines de spécialisation et leur nombre d'années d'expérience en enseignement.

Le chapitre 5 du *Cadre d'évaluation du PPCE 2016* donne de plus amples détails au sujet du cadre conceptuel du questionnaire du PPCE. Ce cadre conceptuel reflète les conclusions des recherches et pratiques exemplaires actuelles dans le domaine du développement de la littératie et de l'apprentissage de la lecture. Il comprend également des renseignements issus des questionnaires (destinés aux élèves, au personnel enseignant et aux écoles) permettant de recueillir des données contextuelles.

Mise à jour du cadre des questionnaires

Il a été demandé aux spécialistes de la lecture d'examiner et de réviser, indépendamment les uns des autres, le cadre des questionnaires utilisé pour le PPCE 2007, l'année précédente où la lecture était le domaine principal de l'évaluation. Il leur a été demandé de cerner les importants aspects du contexte dans lequel les élèves apprennent à lire et deviennent alphabètes, ainsi que de cerner les domaines qui pourraient être améliorés ou les sujets qui pourraient être ajoutés selon la documentation universitaire sur la littératie, qui fournit un contexte à la manière dont les élèves apprennent à lire et aux défis auxquels ils font face.

Le but de l'examen était d'élaborer trois questionnaires concis se concentrant sur des questions relatives à l'apprentissage et à l'enseignement de la lecture, soit le domaine principal, et qui pourraient fournir d'importants renseignements contextuels aux provinces et aux territoires. Ces questionnaires axés sur la lecture et mis au point par les spécialistes ont été traduits et

révisés par le CMEC, puis envoyés aux membres du Comité des évaluations pancanadiennes et internationales pour qu'ils les examinent et apportent des révisions supplémentaires.

L'évaluation du PPCE 2016 comprenait trois questionnaires : un questionnaire pour les élèves participant à l'évaluation, un questionnaire pour leurs enseignantes et enseignants des domaines langagiers de 8^e année/2^e secondaire et un questionnaire pour les directions d'école. Ces questionnaires portaient également sur le besoin particulier de cerner les facteurs associés au rendement en lecture et avaient pour but de fournir un contexte aux résultats de l'évaluation. Ils comprenaient des données descriptives de base utiles à la fois à l'élaboration de politiques et à la recherche, comme le statut socio-économique des élèves, les caractéristiques démographiques de l'école et les qualifications du personnel enseignant. Divers sujets permettaient de cerner les éléments utiles à l'élaboration de politiques. Les questions mettaient principalement l'accent sur la lecture, domaine principal de l'évaluation, mais incluaient également des items portant sur les stratégies d'enseignement et d'apprentissage et les comportements. D'autres questions étaient conformes aux orientations mises de l'avant par les ministères de l'Éducation, même si ces dernières n'ont pas de lien évident avec le rendement dans le domaine principal. L'objectif de cette sélection de thèmes était de fournir des renseignements utiles à la recherche en lecture.

Identité sexuelle

L'éducation inclusive est jugée importante dans les provinces et les territoires du Canada, et a donné lieu à l'élaboration de politiques et de ressources visant à appuyer l'inclusion. Un aspect de l'éducation inclusive est l'identité sexuelle. Dans les questionnaires destinés aux élèves, au personnel enseignant et à l'école du PPCE 2016, la question du genre a été élargie pour inclure deux choix supplémentaires pour les personnes interrogées, comme le montre l'encadré ci-dessous.

<p>Comment t'identifies-tu?</p> <ul style="list-style-type: none"><input type="radio"/> Garçon<input type="radio"/> Fille<input type="radio"/> Je m'identifie d'une autre façon.<input type="radio"/> Je préfère ne pas le dire.

Différences entre les sexes

Les différences de rendement en lecture en faveur des élèves de sexe féminin sont un élément constant dans les évaluations à grande échelle, tant au niveau pancanadien qu'au niveau international. Il paraissait donc important de trouver les causes potentielles du phénomène observé dans les questionnaires, en se concentrant explicitement sur les aspects suivants :

- traitement différentiel des garçons et des filles à l'école;
- comportements ou champs d'intérêt différents en ce qui concerne la lecture hors de l'école.

Même si, pour ce qui est des mathématiques et des sciences, cet aspect est moins souligné, il sera tout de même intéressant de suivre les tendances de l'écart entre les sexes au fil du temps.

Confidentialité

Les questionnaires du personnel enseignant et de l'école étaient reliés aux résultats des élèves, mais portaient un numéro d'identification unique afin de préserver la confidentialité.

Chapitre 4. Procédures d'échantillonnage

Au printemps 2016, la quatrième évaluation du PPCE a été réalisée. Trois domaines ont été évalués : la lecture, les mathématiques et les sciences – la lecture étant le domaine principal. Quatre livrets d'évaluation ont été utilisés pour l'ensemble des domaines de l'évaluation, la majorité des items portant sur la lecture. Un seul niveau scolaire (8^e année/2^e secondaire) a été évalué. Dix-huit populations ont été associées à la démarche.

Ce chapitre présente le plan d'échantillonnage pour cette évaluation et explique le déroulement des activités reliées à la sélection des échantillons.

Plan d'échantillonnage

Entre 1993 et 2004, le CMEC a été associé à des évaluations pancanadiennes par l'entremise du PIRS. En 2007, le PPCE a remplacé le PIRS. Ce programme d'évaluation a conservé certaines caractéristiques des évaluations du PIRS, mais certains aspects techniques ont été modifiés : trois disciplines sont évaluées à chaque cycle, l'une des disciplines étant considérée comme « principale », les deux autres comme « secondaires ». Plusieurs livrets d'évaluation sont utilisés. Depuis 2010, la population à évaluer a été définie par rapport à un niveau d'enseignement plutôt qu'en fonction de l'âge. L'exploitation des données recueillies sur le rendement privilégie deux axes : le calcul de niveaux de rendement pour le domaine principal et la compilation de résultats moyens pour tous les domaines évalués. Le plan d'échantillonnage a dû être adapté pour s'assurer que les données recueillies répondront aux besoins analytiques et pour être généralisables à l'échelon pancanadien.

Un critère majeur de l'échantillonnage pour l'étude principale – outre la représentation de la population canadienne – est l'uniformité des procédures d'un cycle à l'autre. Pour l'étude principale du PPCE 2016, le CMEC a retenu les services et travaillé en étroite collaboration avec Statistique Canada en ce qui concerne le plan d'échantillonnage, la mise en œuvre de l'échantillonnage et l'attribution de pondérations aux données recueillies. Le présent chapitre résume le plan d'échantillonnage du PPCE 2016; des éléments plus détaillés concernant l'échantillonnage et la pondération se trouvent dans le rapport préparé par Statistique Canada (Annexe A).

Définir la population à partir de laquelle un échantillon est sélectionné est une étape essentielle de l'élaboration d'un plan d'échantillonnage solide. Une bonne définition de la population cible⁴ facilite le processus d'échantillonnage et évite les ambiguïtés. Le tableau 4.1 montre la population représentée dans le plan d'échantillonnage du PPCE. Pour valider l'exactitude de la base de sondage, Statistique Canada a comparé cette base aux projections démographiques du recensement canadien pour les personnes de 13 ans, soit l'âge de la majorité des élèves de 8^e année/2^e secondaire. Il a été conclu que cette base était bien harmonisée avec les

⁴ Par « population cible », on entend ici les écoles admissibles à la sélection après l'exclusion de celles ne répondant pas aux critères retenus par le CMEC ou les provinces/territoires concernés. La « population globale », pour sa part, est formée de toutes les écoles accueillant des élèves de 8^e année/2^e secondaire.

projections démographiques de Statistique Canada. Ces statistiques sont tirées des données que les provinces avaient fournies au CMEC pour la mise à l'essai de l'évaluation en 2015.

Tableau **Error! No text of specified style in document..2** – Couverture de la population cible du PPCE

PROVINCE	Projections démographiques* (personnes de 13 ans)		Base de sondage**		Proportion de la population (%)	
	2013	2015	2013	2015	2013	2015
Colombie-Britannique	47 129	45 848	46 895	45 576	99,5	99,4
Alberta	45 386	45 354	40 094	46 578	88,3	102,7
Saskatchewan	13 677	13 371	12 598	12 335	92,1	92,3
Manitoba	16 289	15 852	14 451	14 174	88,7	89,4
Ontario	153 091	147 322	145 756	146 765	95,2	99,6
Québec	79 921	78 354	85 435	81 563	106,9	104,1
Nouveau-Brunswick	8 010	7 543	7 976	7 327	99,6	97,1
Nouvelle-Écosse	9 785	9 146	9 792	8 674	100,1	94,8
Île-du-Prince-Édouard	1 657	1 527	1 487	1 433	89,7	93,8
Terre-Neuve-et-Labrador	5 436	5 221	5 441	5 193	100,1	99,5
CANADA	380 381	369 538	369 925	369 618	97,3	100

*Source : Statistique Canada. Tableau 051-0001 – Avec données tirées des estimations de la population au 1^{er} juillet, par groupe d'âge et par sexe, Canada, provinces et territoires, fréquence annuelle (sauf indication contraire)

** La base a été produite à partir de la liste d'écoles fournie par les provinces au CMEC.

Base de sondage

Base de sondage de premier degré – liste des écoles admissibles

Comme pour les évaluations du PIRS et les précédents cycles du PPCE, une procédure d'échantillonnage à deux degrés a été utilisée pour le PPCE 2016 : dans un premier temps, la sélection d'écoles participantes (base de sondage de premier degré – liste des écoles admissibles) et, dans un deuxième temps, le choix d'une classe de 8^e année/2^e secondaire dans les écoles retenues (base de sondage deuxième degré – liste des classes admissibles). Compte tenu de la taille des populations visées par l'évaluation, un recensement des écoles de certains groupes cibles a eu lieu et des élèves ont pu être sélectionnés dans ces écoles. Dans certains cas, il y a eu un recensement des élèves de 8^e année/2^e secondaire. Pour être en mesure de généraliser les statistiques produites pour les élèves faisant partie d'un échantillon, celui-ci doit respecter plusieurs critères⁵. Il s'agit notamment de critères relatifs à la taille de l'échantillon, à l'exclusion et à l'inclusion *a priori* d'écoles et au processus employé pour réaliser les sélections.

⁵ Dans le cas du recensement des élèves, il n'y a pas d'inférence statistique; les marges d'erreur n'ont habituellement pas besoin d'être compilées.

Dans la préparation de la base de sondage, certaines écoles ont été exclues en fonction des catégories *a priori* suivantes :

- les écoles spécialisées dans lesquelles tous les élèves avaient des besoins éducatifs spéciaux;
- les écoles au sein d'une autre province;
- les écoles géographiquement isolées;
- les écoles fédérales/internationales;
- les écoles non subventionnées;
- les écoles fermées.

Il convient de noter qu'au premier degré de la base de sondage, les écoles exemptées ont été exclues : elles n'ont donc pas été prises en compte durant le processus d'échantillonnage.

Base de sondage de deuxième degré – liste des classes admissibles

Au deuxième degré de l'échantillonnage, la liste des classes admissibles a été obtenue. Cette base de sondage a été élaborée après réception de toutes les listes des classes des écoles sélectionnées. À ce stade, les élèves ont été exclus en fonction des catégories *a priori* ci-dessous. Une description détaillée de ces catégories se trouve au chapitre 6 et à l'Annexe A. En bref, les critères d'exemption sont les suivants :

- troubles fonctionnels;
- déficiences intellectuelles ou troubles socioaffectifs;
- maîtrise limitée du français ou de l'anglais (locutrices ou locuteurs non natifs).

Ces élèves, ou classes d'élèves, ont été retirés de la base avant l'échantillonnage. Une classe entière peut être exemptée si tous les élèves se trouvent dans une catégorie pour laquelle nous exemptons les élèves. Il a fallu que le CMEC approuve toutes les exemptions de classes.

Stratification

La « stratification » est un moyen d'organiser la base d'échantillonnage afin d'atteindre une meilleure précision avec une taille d'échantillon fixe. Elle peut aussi servir à garantir qu'une taille minimale d'échantillon pour certains groupes de population sera obtenue. Les strates sont exhaustives et constituent des groupes d'écoles qui s'excluent mutuellement, chaque école étant associée à une seule strate. La taille de l'échantillon est répartie entre les strates et chaque strate est échantillonnée de façon indépendante.

Afin de publier des statistiques fiables aux échelons pancanadien et provincial, de même que selon la langue des conseils/commissions/districts scolaires des provinces, un échantillon suffisamment grand était nécessaire dans ces domaines. Ainsi, les strates du PPCE sont définies comme le classement recoupé des provinces selon la langue du conseil, de la commission ou du district scolaire.

Dans certaines provinces, afin que les résultats du PPCE soient représentatifs à la fois de la province et de la langue (population), un recensement des écoles et/ou des élèves a été utilisé pour certaines strates. Le recensement des écoles incluait toutes les écoles ayant des élèves de

8^e année/2^e secondaire comme écoles échantillonnées, et la sélection aléatoire des classes a eu lieu dans ces écoles. Le recensement des élèves comprenait la sélection de tous les élèves de 8^e année/2^e secondaire au sein d'une école sélectionnée. Le tableau 4.2 présente la liste de strates et indique si les écoles ont été échantillonnées au premier degré ou si les recensements ont été utilisés au niveau des écoles et/ou des élèves.

Tableau 4.2 – Répartition de la taille des échantillons du PPCE 2016

Strate	Échantillonnage au premier degré	Recensement des écoles	Recensement des élèves	Nombre attendu d'écoles	Nombre attendu d'élèves
Colombie-Britannique – anglophone	X			150	3 011
Colombie-Britannique – francophone		X		12	171
Alberta – anglophone	X			150	2 990
Alberta – francophone		X		20	299
Saskatchewan – anglophone	X			186	3 143 ⁶
Saskatchewan – francophone		X	X	5	79
Manitoba – anglophone	X			150	2 846
Manitoba – francophone		X		19	267
Ontario – anglophone	X			150	2 713
Ontario – francophone	X			125	2 094
Québec – anglophone		X		89	1 801
Québec – francophone	X			150	3 361
Nouveau-Brunswick – anglophone		X		83	1 516
Nouveau-Brunswick – francophone		X		61	1 164
Nouvelle-Écosse – anglophone		X		116	2 480

⁶ La confusion concernant la classification des écoles à double programme en Saskatchewan (c'est-à-dire des écoles offrant à la fois un programme de langue anglaise et un programme de langue française) a entraîné la sélection d'un échantillon d'élèves anglophones de plus grande taille que ce qui était exigé par le cadre d'échantillonnage. Les écoles de la Saskatchewan à double programme relèvent du système scolaire anglophone.

Nouvelle-Écosse – francophone		X		11	184
Île-du-Prince-Édouard – anglophone		X		22	438
Île-du-Prince-Édouard – francophone		X	X	3	58
Terre-Neuve-et-Labrador – anglophone		X		114	1 865
Terre-Neuve-et-Labrador – francophone		X	X	3	23

Remarque : Bien que les élèves des écoles francophones de l'Île-du-Prince-Édouard aient été échantillonnés, les résultats n'ont pas été présentés en raison de la petite taille de l'échantillon. Même si les élèves des écoles francophones de Terre-Neuve-et-Labrador ont été pris en considération dans la répartition de l'échantillon, ces élèves n'ont pas participé au PPCE 2016.

Taille des échantillons

La taille d'un échantillon est reliée à l'importance numérique de la population, à la marge d'erreur et au niveau de confiance que l'on est prêt à accepter au moment des compilations statistiques pour être en mesure de généraliser les données pour les populations évaluées.

L'utilisation de plusieurs livrets d'évaluation et le regroupement d'élèves par niveaux de rendement ont un effet direct sur la taille des échantillons. En effet, en tenant compte de ces deux paramètres, les marges d'erreur subiraient des variations considérables. Un nombre suffisamment important d'élèves a donc été sélectionné afin de garantir une marge d'erreur d'au maximum 3 p. 100 en général, avec un niveau de confiance de 95 p. 100, ce qui était conforme aux niveaux des précédents cycles du PPCE. Le tableau 4.3 indique la formule utilisée pour déterminer la taille d'un échantillon par rapport au calcul des distributions de fréquences.

Tableau 4.3 – Estimation de la taille d'un échantillon

$$n = \frac{Nz^2pq}{Nd^2 + z^2pq}$$

N = taille de la population

z = valeur de l'abscisse sur la courbe normale correspondant au niveau de confiance désiré

p = proportion observée dans l'échantillon

q = 1 – p

d = précision désirée, c'est-à-dire marge d'erreur que l'on est prêt à accepter

Échantillonnage au premier degré – échantillonnage des écoles

Bases de données sur les écoles

Dans le but de réaliser les travaux reliés à l'échantillonnage, le CMEC devait mettre au point des bases de données sur chaque population évaluée. Chaque province devait utiliser un même fichier préparé par le CMEC afin de dresser la liste des écoles et de préparer les autres informations nécessaires. Les variables requises pour chaque école incluaient le nombre total d'élèves dans chaque école, le nombre d'élèves de 8^e année/2^e secondaire, la langue du conseil, de la commission ou du district scolaire, l'information précisant si l'établissement était une école d'immersion, ainsi que d'autres renseignements sur les écoles.

Sélection des écoles

La sélection des écoles a été effectuée par Statistique Canada et deux méthodes ont été utilisées à ce degré : l'utilisation des recensements susmentionnés et l'utilisation d'un échantillonnage systématique pour les strates sans recensement. L'échantillonnage systématique est examiné plus en détail dans le document sur l'échantillonnage qui se trouve à l'Annexe A. C'est à ce degré que les écoles échantillonnées et les écoles de remplacement sont sélectionnées.

Exclusion d'écoles

La décision touchant l'exclusion de certaines catégories d'écoles ou de certaines écoles a été prise par chaque coordonnatrice ou coordonnateur provincial ou territorial. Cependant, le nombre d'élèves touchés par ces exclusions ne devait pas dépasser une certaine proportion (environ 2 p. 100) de la population totale. Il faut remarquer que les écoles exclues dans l'échantillonnage devaient toutefois apparaître dans les fichiers présentant les données sur la population à évaluer.

Le CMEC a recueilli les informations statistiques sur les écoles de chaque population, en utilisant les paramètres présents dans les fichiers sur les écoles préparés par les provinces. Ces informations incluaient :

- le nombre d'écoles et d'élèves dans la population totale;
- le nombre d'écoles et d'élèves exclus de la population totale;
- le nombre d'écoles et d'élèves faisant partie de la population cible (c'est-à-dire la population totale moins les exclusions);
- après la sélection des écoles, le nombre d'écoles et d'élèves faisant partie de l'échantillon sélectionné.

Si les données indiquaient que le critère retenu pour les exclusions n'avait pas été respecté (2 p. 100 ou moins des élèves exclus *a priori*), le CMEC a communiqué avec les provinces concernées.

Il était très important que la proportion d'élèves concernés par l'exclusion de certaines écoles respecte les critères établis. Plusieurs motifs peuvent justifier l'exclusion *a priori* de certaines écoles : taille, distance, clientèle spéciale ou école relevant d'une province autre que celle où elle est située. Les coordonnatrices et coordonnateurs devaient indiquer au CMEC les numéros d'identification des écoles à exclure, ainsi que les motifs justifiant cette décision.

Ces informations étaient codifiées dans la strate prévue à cette fin. Il faut noter que toutes les écoles devaient être incluses dans les fichiers de données sur chaque population à évaluer. Il était en effet nécessaire de connaître, pour chaque population évaluée, l'effectif total des élèves de 8^e année/2^e secondaire.

Échantillonnage au deuxième degré – échantillonnage des élèves

Comme nous l'avons indiqué plus haut, l'échantillonnage pour l'évaluation du PPCE a été effectué en deux étapes. Dans un premier temps, des écoles ont été choisies lorsqu'il ne s'agissait pas d'un recensement des écoles devant participer à cette évaluation. Il était cependant possible de ne pas soumettre tous les élèves de 8^e année/2^e secondaire fréquentant une école sélectionnée à l'évaluation du PPCE. Il a donc fallu que Statistique Canada échantillonne les élèves qui participaient à l'activité. Cette sélection devait respecter des règles rigoureuses pour que les échantillons d'élèves représentent les populations évaluées. C'est le CMEC qui a choisi de façon aléatoire la classe de 8^e année/2^e secondaire des écoles sélectionnées pour participer à l'évaluation. Voici la démarche qui a été retenue pour assurer la réalisation de la sélection des élèves.

1. Dans un premier temps, chaque coordonnatrice ou coordonnateur des provinces a présenté la liste de toutes les écoles admissibles avec des élèves de 8^e année/2^e secondaire qui relevaient de l'autorité de sa province respective.
2. Le CMEC a sélectionné les écoles devant participer au PPCE et envoyé la *Liste des écoles* aux coordonnatrices et coordonnateurs des provinces.
3. Les coordonnatrices et coordonnateurs ont communiqué avec les écoles sélectionnées et demandé la liste des classes de 8^e année/2^e secondaire. Cette liste a été transmise au CMEC.
4. Le CMEC a sélectionné les classes devant participer au PPCE et envoyé la liste des classes aux coordonnatrices et coordonnateurs des provinces. Il était possible, dans certains cas, que plus d'une classe soit choisie dans la même école. Après consultation des écoles, les coordonnatrices et coordonnateurs des provinces pouvaient choisir de retirer telle ou telle classe de la participation au PPCE. Dans ce cas, il leur fallait communiquer avec le CMEC pour assurer la sélection d'une classe de substitution. Il fallait que les coordonnatrices et coordonnateurs des provinces comprennent que ce type de substitution ne pouvait se faire que dans des circonstances exceptionnelles et devait être approuvé par le CMEC.
5. Les coordonnatrices et coordonnateurs ont demandé aux écoles sélectionnées de dresser la *Liste des élèves* pour chaque classe de 8^e année/2^e secondaire sélectionnée pour la participation au PPCE. Ces listes indiquaient également les noms des élèves dans l'incapacité de participer au PPCE et mentionnaient tout besoin particulier. Il était

demandé aux directions d'école de dresser la liste de tous les élèves de 8^e année/2^e secondaire comme suit :

- i. dans la mesure du possible, la liste de tous les groupements de classe de 8^e année/2^e secondaire (par exemple, 8A, 8B) ayant lieu lors de la première période du premier jour du cycle scolaire normal (soit un cycle de cinq jours ou de sept jours) — c'était l'option A;
 - ii. si l'option A n'était pas possible, alors la *Liste des élèves* actuellement inscrits en 8^e année/2^e secondaire, par ordre alphabétique.
6. Après l'évaluation, les coordonnatrices et coordonnateurs des provinces ont communiqué au CMEC la *Liste des élèves* ayant participé au PPCE de 2016. Les listes préparées à l'étape 5 ont été utilisées. Il a fallu inscrire les raisons motivant la non-participation à l'évaluation de certains élèves.

Tableau 4.4 – Résumé des critères utilisés pour échantillonner les élèves, selon les types de strates

Type de strate	Méthodes d'échantillonnage
Strate échantillonnée à deux niveaux	(a) Écoles ayant plus de 20 élèves : échantillonnage aléatoire simple – une classe de 8 ^e année/2 ^e secondaire; environ un quart des élèves sera évalué à l'aide de l'un des quatre livrets. (b) Écoles ayant plus de 20 élèves, mais qui ne sont pas en mesure de dénombrer les classes de 8 ^e année/2 ^e secondaire : le dénombrement de tous les élèves de 8 ^e année/2 ^e secondaire sera effectué et un échantillonnage aléatoire simple de 20 élèves sera sélectionné.
Recensement des écoles	(c) Écoles ayant 20 élèves ou moins : recensement des élèves. Une classe de 8 ^e année/2 ^e secondaire est échantillonnée pour toutes les écoles figurant sur la liste; un quart des élèves de chaque classe sera évalué à l'aide de l'un des quatre livrets.
Recensement des élèves	Tous les élèves figurant sur la liste sont échantillonnés; un quart des élèves de chaque classe sera évalué à l'aide de l'un des quatre livrets.

Le processus d'échantillonnage constitue un volet très important dans une activité d'évaluation comme le PPCE. Il en va souvent de la crédibilité des résultats qui seront rendus publics au terme du projet. Le choix des écoles invitées à participer au PPCE est réalisé de façon centralisée à partir des renseignements fournis par les coordonnatrices et coordonnateurs

provinciaux. Quant au choix des élèves devant être évalués, c'est le CMEC qui sélectionne la classe ou les classes de chaque école faisant partie des échantillons choisis.

Pondération de l'échantillon

Après l'achèvement de la collecte de données, les fichiers de données contenant les informations sur la participation ont été envoyés à Statistique Canada afin de calculer les pondérations. Des pondérations ont été attribuées aux élèves, au personnel enseignant et aux écoles. Les détails concernant les procédures de pondération se trouvent à l'Annexe A.

Chapitre 5. Mise à l'essai

Les items présentés aux élèves dans les évaluations à grande échelle doivent être vérifiés quant à leur qualité intrinsèque, de même qu'en matière d'applicabilité à la population cible. Les items élaborés par des spécialistes de contenu sont testés à cette étape du processus. La mise à l'essai comporte un plus grand nombre d'items que l'étude principale, afin que seuls les meilleurs soient retenus pour évaluer le rendement des élèves de 8^e année/2^e secondaire.

Groupe de travail sur l'examen des items — Mise à l'essai

Le groupe de travail sur l'examen des items, composé de représentantes et représentants de quatre provinces, s'est réuni pour examiner le contenu, le vocabulaire et la traduction de tous les items de lecture, et pour vérifier la classification des items selon les sous-domaines du test. Ce groupe a également examiné les items afin de mettre en évidence tout problème concernant le niveau de vocabulaire ou les biais (relatifs au sexe, à la culture, à la géographie, etc.). Le groupe de travail a sélectionné des items pour la mise à l'essai qui couvraient bien les différents sous-domaines en lecture, avec différents niveaux de difficulté.

Livrets d'évaluation

Trois livrets d'évaluation ont été préparés pour la mise à l'essai. Chaque livret était conforme aux spécifications figurant dans le *Cadre d'évaluation du PPCE 2016* et contenait environ 40 items en lecture, ainsi que le questionnaire de l'élève. Les élèves avaient 90 minutes pour remplir le livret et 30 minutes pour remplir le questionnaire. Les questionnaires du personnel enseignant et de l'école ont été préparés sous la forme de livrets séparés.

Séance de correction

La séance de correction s'est déroulée sur cinq jours à Ottawa, en juillet 2015. Approximativement 2000 livrets ont été corrigés, dont environ 1000 livrets en français et 1000 livrets en anglais. La séance comportait deux chefs de table, dont l'un pour la table francophone et l'autre pour la table anglophone, ainsi que 20 codeuses et codeurs, dont la moitié se trouvait à la table francophone et l'autre moitié à la table anglophone.

Le processus de codage a employé des tests de fiabilité croisés, deux fois par jour, pour s'assurer que les correctrices et correcteurs évaluaient les items de la même façon et selon les codes attribués par les spécialistes. Le degré d'uniformité entre les correctrices et correcteurs et les spécialistes était généralement supérieur à 85 p. 100. Pour les quelques items où le test de fiabilité donnait des résultats moins uniformes, les correctrices et correcteurs ont réexaminé les documents de formation et procédé ensuite à une nouvelle correction des items.

Saisie des données

Les élèves ont indiqué leurs réponses aux items à réponse choisie (par exemple, réponse à choix multiples, vrai ou faux) sur un feuillet de réponses détachable et ont écrit leurs réponses aux items à réponse construite directement dans les livrets d'évaluation du PPCE. Pour les

items à réponse choisie, chaque réponse correcte recevait un point. Pour les items à réponse construite, qui pouvaient recevoir un plein crédit ou un crédit partiel, le score attribué se situait sur un barème de 0 à 3.

Lors de la séance de correction pour la mise à l'essai, les correctrices et correcteurs n'ont codé que les items à réponse construite en noircissant dans un feuillet de correction le cercle correspondant le mieux à la réponse de l'élève. Les correctrices et correcteurs avaient reçu chacun un paquet de 10 livrets d'évaluation qui provenaient de différentes provinces. Une fois la séance de correction terminée, tous les livrets d'évaluation ont été envoyés à une entreprise de saisie de données, où les données des items à réponse choisie et des questionnaires de l'élève, de l'école et du personnel enseignant et les codes des correctrices et correcteurs pour les items à réponse construite ont été capturés. Les données recueillies ont été fusionnées dans un fichier Excel pour créer une base de données.

Analyse des données

Les données de la mise à l'essai pour les items de lecture ont été analysées simultanément par un spécialiste de la psychométrie du CMEC et par un spécialiste externe. Ce processus parallèle avait pour but de garantir la validation des données. Les données ont été présentées au comité consultatif technique du PPCE, qui a examiné l'analyse, les bases de données, les fichiers et les règles suivies pour la saisie des données (par exemple, pour la pondération des items).

Une seule base de données de mise à l'essai a été créée pour les items de l'évaluation et des questionnaires. En plus d'un score de rendement global, des scores pour chaque scénario et sous-domaine ont été produits.

L'analyse des données s'est effectuée selon la théorie classique. Le comité consultatif technique du PPCE a utilisé les données ainsi obtenues pour sélectionner, d'un point de vue statistique, les meilleurs items pour l'étude principale et pour rejeter les items aberrants, c'est-à-dire ceux qui ne se comportaient pas comme les autres items du test. Pour analyser les items, des indices statistiques ont été utilisés, dont un indice de difficulté et un indice de discrimination, afin de vérifier les qualités psychométriques de chaque item. L'indice de difficulté est déterminé à partir de la valeur p , p étant la proportion des individus qui ont répondu correctement à un item par rapport au nombre total d'individus qui ont répondu à cet item. Les spécialistes ont également vérifié la discrimination des items, pour veiller à ce que chaque item distingue les élèves plus forts des élèves plus faibles. Le coefficient alpha de Cronbach a aussi été utilisé pour estimer la cohérence interne du test.

Les spécialistes en statistiques ont également procédé à d'autres analyses potentiellement pertinentes, comme le calcul des moyennes pour chaque item et la préparation des distributions des fréquences pour le pourcentage d'élèves ayant sélectionné chaque réponse pour les items à réponse choisie ou ayant obtenu chaque code pour les items à réponse construite. Ils ont également analysé le pourcentage de données manquantes et effectué une analyse du fonctionnement différentiel des items (FDI) selon le sexe et la langue.

Groupe de travail sur la sélection des items – Étude principale

Après la mise à l'essai des items, le groupe de travail sur la sélection des items, composé de représentantes et représentants de sept provinces, s'est réuni pour examiner et sélectionner des passages et des items pour l'étude principale. Tous les livrets d'évaluation ont été fournis aux membres du groupe, de même que les résultats et les statistiques pour chaque item, afin qu'ils puissent vérifier la qualité, le degré de difficulté et l'équivalence du fonctionnement des items dans les deux langues et pour tous les sexes. Ils ont aussi pris en compte les commentaires des codeuses et codeurs qui avaient participé à la séance de correction des items de la mise à l'essai (à partir du questionnaire qui leur avait été donné à la fin de la séance de correction). Ce questionnaire contenait des remarques pertinentes sur l'instrument d'évaluation en général, de même que des commentaires sur chaque item concernant sa qualité.

Le groupe de travail a sélectionné les items pour les trois domaines : lecture, mathématiques et sciences. Tous les items de lecture, de mathématiques et de sciences qui devaient être utilisés comme items d'ancrage étaient identiques à ceux utilisés lors de l'évaluation où ils faisaient partie du domaine principal : lecture pour le PPCE de 2007, mathématiques pour le PPCE de 2010 et sciences pour le PPCE de 2013. Comme il ne fallait qu'un petit sous-ensemble d'items, le groupe de travail a pris soin de bien représenter chaque sous-domaine et un éventail de niveaux de difficulté.

Le groupe de travail comprenait des spécialistes bilingues qui devaient comparer les versions françaises et anglaises des livrets d'évaluation, afin de déterminer si les élèves réussissaient mieux pour un item dans une langue que dans l'autre et, le cas échéant, de discuter des raisons possibles de tels résultats.

Le groupe de travail a également examiné les réponses des questionnaires de l'élève, de l'école et du personnel enseignant au niveau du contenu et d'un point de vue statistique et psychométrique. Il a choisi les items qui, selon lui, pouvaient fournir le plus de données pertinentes lors de l'étude principale, comme l'établissement de liens entre les données contextuelles et le rendement des élèves.

Révision du cadre d'évaluation

La mise à l'essai des items a fourni des informations sur les items qui ont facilité la sélection des meilleurs items pour l'étude principale. Le *Cadre d'évaluation du PPCE 2016* a été par la suite réexaminé afin de s'assurer qu'il correspondait bien aux items de l'évaluation. Très peu de changements ont été nécessaires.

Chapitre 6. Étude principale

L'évaluation du PPCE a eu lieu entre le 25 avril et le 20 mai 2016⁷, le domaine principal étant la lecture et les domaines secondaires étant les mathématiques et les sciences. Plus de 30 000 élèves, sélectionnés de façon aléatoire dans plus de 1500 écoles dans 10 provinces du Canada, ont participé à l'évaluation en français et en anglais. Les principaux items ont permis d'évaluer les connaissances et les habiletés des élèves de 8^e année/2^e secondaire dans les trois matières.

Livrets d'évaluation

Chaque livret d'évaluation contenait deux blocs d'items en lecture, un bloc d'items en mathématiques et un bloc d'items en sciences. Pour évaluer l'équivalence de chaque livret, un sous-ensemble d'items de chaque domaine était répété dans des paires de livrets. Lors de la mise en page des livrets, le CMEC a pris soin d'inclure des scénarios et des items sélectionnés par les groupes de travail et de les présenter de la même manière dans les deux langues.

Réexamen du matériel d'évaluation

Avant de finaliser le matériel d'évaluation, tous les coordonnateurs et coordonnatrices des provinces ont dû les examiner, de façon à ce que leurs commentaires puissent être incorporés le cas échéant. Le matériel envoyé aux provinces à des fins de réexamen comprenait toutes les versions des livrets d'évaluation, le questionnaire de l'élève, le questionnaire de l'école, le questionnaire du personnel enseignant et les documents administratifs connexes. Le CMEC a reçu l'approbation de chaque province pour le matériel d'évaluation.

Impression des livrets d'évaluation

Une fois l'approbation de toutes les provinces reçue, les livrets d'évaluation ont été imprimés à des fins de vérification, pour s'assurer que toutes les modifications de contenu apportées par les groupes de travail et le CMEC avaient été intégrées aux nouvelles versions des livrets d'évaluation. Une fois ce processus terminé, les épreuves fournies par l'imprimeur ont été examinées et approuvées, après quoi les livrets d'évaluation ont été convertis en format PDF et imprimés. Un numéro d'identification unique muni d'un code à barres a été imprimé sur chaque livret afin de l'attribuer à l'élève correspondant. Les livrets d'évaluation et les documents administratifs ont ensuite été emballés pour chaque école et, en fonction des consignes de chaque province, envoyés soit aux coordonnatrices et coordonnateurs des provinces, qui devaient les distribuer aux écoles sélectionnées, soit directement à la coordonnatrice ou au coordonnateur de l'école.

⁷ Les évaluations ont eu lieu à des dates différentes en Saskatchewan, au Québec et à l'Île-du-Prince-Édouard.

Vérification des documents

Chaque coordonnatrice ou coordonnateur de l'école ou de la province devait s'assurer qu'elle ou il disposait des documents destinés à chaque école. Tout colis manquant devait être immédiatement signalé au CMEC, afin de s'assurer que les documents seraient bien arrivés avant la date prévue de l'évaluation. Si les directions d'école et les conseils, commissions ou districts scolaires avaient des questions ou avaient besoin de plus amples informations sur l'évaluation ou le matériel d'évaluation, ils étaient invités à communiquer avec la coordonnatrice ou le coordonnateur de leur province directement.

Lettre envoyée aux parents/tutrices ou tuteurs des élèves

Avant l'étude principale, la coordonnatrice ou le coordonnateur de l'école devait informer les élèves sélectionnés ainsi que leurs parents/tutrices ou tuteurs. Une brochure a été distribuée aux parents pour les informer de l'objectif et de l'importance de l'évaluation.

Procédures administratives

Chaque école sélectionnée devait nommer une coordonnatrice ou un coordonnateur chargé du déroulement de l'évaluation du PPCE dans l'établissement. L'évaluation devait ensuite se dérouler conformément aux procédures établies par le CMEC, pour garantir l'uniformité du déroulement dans toutes les écoles sélectionnées. Avant de procéder à l'évaluation, la coordonnatrice ou le coordonnateur devait se familiariser avec les documents administratifs, en particulier avec le *Guide d'information à l'intention de l'école*, qui décrivait les procédures administratives pour le test. Si la coordonnatrice ou le coordonnateur avait des questions sur l'évaluation, cette personne devait communiquer avec la coordonnatrice ou le coordonnateur de la province.

Chaque élève avait un numéro d'identification unique imprimé sur la couverture de son livret d'évaluation, sur le feuillet de réponse détachable et sur la feuille de correction. Les numéros d'identification ont été attribués afin de préserver la confidentialité des élèves. Les noms des élèves dans la *Liste des élèves* ont été utilisés pour faciliter le déroulement de l'évaluation dans les écoles. Si un élève avait rejoint la classe sélectionnée après l'échantillonnage, il était autorisé à participer à l'évaluation et devait utiliser l'un des livrets supplémentaires fournis.

La *Liste des élèves* indiquait les élèves ayant des besoins spéciaux. Le CMEC fournit aux écoles le matériel d'évaluation nécessaire pour que ces élèves puissent participer à l'évaluation sans en compromettre l'intégrité. Par exemple, même si l'évaluation ne pouvait être mise à disposition dans un format électronique pour les élèves ayant une déficience visuelle, il était possible de produire des versions en gros caractères pour répondre à leurs besoins.

Lorsqu'un élève sélectionné ne pouvait participer à l'évaluation pour une raison quelconque, la coordonnatrice ou le coordonnateur de l'école n'était en aucun cas autorisé à remplacer cet élève; l'élève sélectionné devait au contraire être exempté de l'évaluation et cela devait être indiqué dans le *Formulaire de suivi des élèves*.

Lieu de l'évaluation

Les coordonnatrices et coordonnateurs des écoles devaient trouver un lieu pour le déroulement de l'évaluation du PPCE. Il était essentiel de choisir un endroit calme où les élèves disposaient d'un espace de travail suffisant pour pouvoir répondre aux questions de l'évaluation sans interruption. Dans la mesure du possible, il était conseillé d'organiser l'évaluation le matin afin d'obtenir le meilleur rendement possible de la part des élèves.

Déroulement de l'évaluation

Au début de l'évaluation, la coordonnatrice ou le coordonnateur de l'école a distribué un exemplaire du livret d'évaluation à chaque élève indiqué dans le *Formulaire de suivi des élèves*. Les quatre livrets ont été distribués de façon égale aux élèves de la classe. La coordonnatrice ou le coordonnateur devait en outre s'assurer d'avoir donné les instructions aux élèves avant le début de l'évaluation; les élèves devaient être informés qu'ils disposaient de 90 minutes pour répondre aux questions de l'évaluation. En cas de besoin, les élèves pouvaient prendre 30 minutes supplémentaires pour terminer l'évaluation. Ils disposaient également de 30 minutes pour répondre au Questionnaire de l'élève.

Pour chaque élève, la coordonnatrice ou le coordonnateur de l'école devait inscrire un code de participation dans le *Formulaire de suivi des élèves*. Cette procédure devait permettre, par la suite, de comparer la *Liste des élèves* sélectionnés aux livrets d'évaluation et d'établir le statut de l'élève, afin de déterminer si ce dernier avait participé à l'évaluation, en avait été exempté ou avait été absent.

Une fois l'évaluation terminée, la coordonnatrice ou le coordonnateur devait ramasser tous les documents d'évaluation et les ranger en lieu sûr, afin que le matériel d'évaluation reste confidentiel.

Les coordonnatrices et coordonnateurs des provinces devaient également veiller au bon déroulement de l'évaluation. Ils étaient chargés d'observer le déroulement de l'évaluation dans environ 5 à 10 p. 100 des écoles de leur région. Ils devaient effectuer des suivis au téléphone et mener des observations directes dans les écoles afin de recueillir les renseignements nécessaires sur le plan administratif. S'ils se rendaient dans les écoles à des fins d'observation, ils devaient seulement noter dans quelle mesure les procédures administratives étaient bien suivies et ne pouvaient, en aucun cas, intervenir pendant l'évaluation. Les principaux aspects à observer étaient la sécurité entourant le matériel d'évaluation, le respect des directives données aux écoles, le respect du temps alloué et le respect des règles sur la façon de répondre aux questions des élèves. Ils devaient noter leurs observations dans le Rapport d'observation provincial.

Élèves ayant des besoins spéciaux

Pour cette évaluation, les adaptations étaient définies comme des modifications qui ne compromettaient ni l'intégrité ni le contenu du test, mais qui donnaient à tous les élèves une chance égale de montrer leurs connaissances et leurs compétences au moment de l'évaluation.

Les élèves ayant besoin d'adaptations devaient être identifiés au préalable, dans la liste des élèves admissibles de l'école. Les coordonnatrices et coordonnateurs des écoles devaient aviser la coordonnatrice ou le coordonnateur de leur province dès que possible lorsqu'un élève était identifié comme ayant des besoins spéciaux, afin de garantir que les versions spéciales de l'évaluation soient bien incluses dans l'envoi des livrets d'évaluation à l'école. Il était important de prendre les dispositions nécessaires pour permettre aux élèves ayant des besoins spéciaux de participer autant que possible à l'évaluation sans risquer de compromettre l'intégrité de cette dernière.

Les adaptations étaient autorisées uniquement pour les élèves qui en bénéficiaient habituellement dans le cadre de leur travail régulier en classe. La liste des adaptations autorisées comprenait les éléments suivants : braille, version en gros caractères, papier coloré et version audio. Ces adaptations étaient accordées uniquement aux élèves dont le nom avait été fourni dans la liste des élèves admissibles, en raison du temps supplémentaire nécessaire pour les préparer.

D'autres adaptations étaient également à la disposition de l'ensemble des élèves :

- temps supplémentaire;
- une ou plusieurs pauses durant lesquelles les élèves restaient sous surveillance (la durée d'évaluation n'inclut pas les pauses).

Les coordonnatrices et coordonnateurs des écoles ne pouvaient en aucun cas aider les élèves à interpréter les documents fournis ni les guider dans leurs réponses. Toute modification ou irrégularité par rapport aux lignes directrices administratives devait être décrite dans le *Rapport de la coordonnatrice ou du coordonnateur de l'école*.

Questionnaires pour la direction de l'école et pour le personnel enseignant

Le *Questionnaire de l'école* était généralement rempli par la directrice ou le directeur de l'école. Les enseignantes et enseignants de sciences des classes sélectionnées devaient remplir le *Questionnaire du personnel enseignant*. Les deux questionnaires étaient offerts en versions imprimée et en ligne. Dans certaines provinces, quelques écoles étaient structurées de telle sorte que les élèves n'étaient pas inscrits ou affectés à un niveau scolaire particulier. Dans un tel cas, tous les enseignants et enseignantes de langues associés aux élèves sélectionnés étaient priés de remplir le questionnaire (à raison d'un questionnaire par enseignante ou enseignant). Chaque questionnaire comportait un numéro d'identification inscrit sur sa page couverture. Un numéro d'identification du questionnaire du personnel enseignant était attribué à chaque enseignante ou enseignant identifié et imprimé sur la page couverture du questionnaire. Les questionnaires devaient être distribués au moment de l'évaluation.

Tous les questionnaires (ou toutes les pages de couverture des questionnaires remplis en ligne) ont été ramassés à la fin de la séance d'évaluation. La coordonnatrice ou le coordonnateur de l'école ne pouvait en aucun cas révéler les noms du personnel enseignant. Après l'évaluation, la coordonnatrice ou le coordonnateur avait l'obligation de détruire toute liste contenant les noms des enseignantes et enseignants, afin de garantir la confidentialité des informations.

Les questionnaires pour la direction de l'école et pour le personnel enseignant visaient à établir des liens entre les réponses aux questionnaires et le rendement des élèves. Les données obtenues ont également fourni d'importants renseignements aux responsables de l'élaboration des politiques. L'utilisation des noms des enseignantes et enseignants avait uniquement pour but de faire le lien entre le numéro d'identification d'un questionnaire de l'élève et celui d'un questionnaire du personnel enseignant.

Participation des élèves à l'évaluation et exemptions

Il était attendu des élèves de 8^e année/2^e secondaire qu'ils aient les capacités nécessaires pour effectuer l'évaluation. Il était donc important que les écoles encouragent vivement les élèves à y participer. Les enseignantes ou enseignants pouvaient utiliser diverses stratégies pour motiver les élèves, mais ils devaient suivre et respecter les procédures administratives lors de l'évaluation.

Il se pouvait, cependant, que certains élèves éprouvent des difficultés ou une grande frustration lors de la participation à l'évaluation. Pour ces élèves, le personnel enseignant pouvait décider à l'avance que l'évaluation ne leur était pas recommandée et donc les exempter de l'épreuve. Par exemple, les élèves de la classe sélectionnée dont les habiletés en lecture, en mathématiques ou en sciences étaient très limitées pouvaient être exemptés par l'école de la participation à l'évaluation. Dans certains cas, l'évaluation risquait de déclencher des réactions émotionnelles ou physiques que le personnel de direction pouvait considérer comme nuisibles à l'élève. Que l'élève ait participé à l'évaluation ou en ait été exempté pour diverses raisons, la coordonnatrice ou le coordonnateur de l'école devait l'indiquer à l'aide des codes de participation du *Guide d'information à l'intention de l'école* et inscrire le code approprié dans le *Formulaire de suivi des élèves*. Il était important d'attribuer un code de participation à tous les élèves sélectionnés, afin de garantir un échantillonnage équitable pour chaque province. Il y avait neuf codes de participation :

- 1 = Absent(e)
- 2 = A participé durant la séance régulière
- 2A = A participé durant la séance régulière avec une adaptation
- 3 = A participé pendant la séance de rattrapage
- 4 = Exempté(e) par l'école
- 5 = Exempté(e), car il était impossible d'apporter les adaptations requises
- 6 = N'est plus inscrit(e) dans cette école/classe
- 7 = Élèves qui ne souhaitent pas participer
- 8 = Aucune donnée disponible

Les trois codes d'exemption sont fournis ci-dessous.

F = Exemption pour cause d'incapacités fonctionnelles. Un élève qui a des incapacités physiques et est incapable de travailler dans le contexte du test du PPCE, même avec l'une des adaptations autorisées, devrait être exempté. Les élèves ayant des incapacités fonctionnelles mais qui sont capables de participer devraient être inclus dans le test. Les sept adaptations autorisées étaient les suivantes :

- même si tous les élèves ont droit à jusqu'à 30 minutes supplémentaires pour terminer l'évaluation, il est possible d'accorder du temps supplémentaire si l'élève bénéficie d'une telle adaptation en situation de test pendant le programme scolaire normal;
- pause ou multiples pauses, à condition que l'élève soit supervisé pendant ces pauses;
- autre local;
- utilisation du braille, de gros caractères d'imprimerie ou de papier en couleur;
- recours à un scribe (cette personne ayant l'obligation d'écrire exactement ce que dit l'élève, mot pour mot, sans modification);
- lecture mot pour mot des instructions uniquement, dans tous les domaines;
- lecture mot pour mot d'instructions ou de questions ponctuelles pour les sections de sciences et de mathématiques seulement (dans les cas où il faut lire en entier le test de sciences ou de mathématiques, une version audio du test sur disque compact peut être fournie).

I = Exemption pour cause d'incapacités intellectuelles ou de problèmes socioaffectifs. Un élève qui est considéré, selon l'opinion professionnelle de la direction de l'école ou d'autres membres qualifiés du personnel, comme ayant des incapacités intellectuelles ou des problèmes socioaffectifs ou dont les tests ont montré qu'il avait de telles incapacités ou problèmes devrait être exempté. Cette catégorie inclut les élèves qui sont incapables, sur le plan mental ou affectif, de suivre ne serait-ce que les instructions générales du test.

N = Exemption en raison de la langue (locutrices ou locuteurs non natifs). Cette exemption ne s'applique qu'aux élèves dont la langue première n'est ni le français ni l'anglais. Dans les évaluations de grande envergure, les écoles peuvent examiner le cas des élèves qui sont au Canada depuis moins de deux ans.

Le nombre et les pourcentages d'élèves exemptés se trouvent au tableau A.2 à l'Annexe A (« Procédures d'échantillonnage et taux de réponse », page 176) du rapport public du PPCE 2016 (O'Grady, Fung, Servage et Khan, 2018)⁸.

Organisation d'une séance de rattrapage

Les coordonnatrices et coordonnateurs des écoles devaient s'assurer que le taux de participation des élèves de leur école était adéquat. Pour ce faire, ils devaient compter le nombre de codes A (absence) et B (participation à la séance prévue) et calculer le taux en pourcentage de participation des élèves à l'évaluation à l'aide de la formule suivante :

$$\frac{(B)}{(A + B)} \times 100$$

⁸ <https://www.cmec.ca/Publications/Lists/Publications/Attachments/381/PCAP-2016-Public-Report-FR.pdf>

Si le taux de participation des élèves était inférieur à 85 p. 100, une séance de rattrapage devait être organisée avant les dates indiquées. Les coordonnatrices et coordonnateurs des écoles devaient tenter d'inclure le maximum d'élèves qui avaient été absents. Si un élève faisait l'évaluation pendant la séance de rattrapage, son code de participation passait de A (absence) à C (participation à la séance de rattrapage) dans le *Formulaire de suivi des élèves*.

Renvoi du matériel d'évaluation

Après l'évaluation des élèves, les coordonnateurs et coordonnatrices des écoles devaient remplir le *Rapport de la coordonnatrice ou du coordonnateur de l'école*. Ils devaient aussi remplir la *Liste des documents expédiés* et indiquer le nombre d'exemplaires de chaque document. Dès que possible après l'évaluation, ils devaient renvoyer au CMEC la *Liste des documents expédiés*, le *Rapport de la coordonnatrice ou du coordonnateur de l'école*, le *Formulaire de suivi des élèves* dûment rempli, tous les questionnaires de l'école et du personnel enseignant, les livrets d'évaluation et les feuilles de réponses, ainsi que les exemplaires et les photocopies des livrets d'évaluation non utilisés.

Séance de correction

La séance de correction des items de l'étude principale a eu lieu à Ottawa (Ontario), en juillet 2016. Tous les items de lecture, de mathématiques et de sciences ont été corrigés par des enseignantes et enseignants de ces domaines. La correction des items de lecture a duré deux semaines, parce qu'il s'agissait du domaine principal et qu'il y avait un plus grand nombre d'items à corriger. La correction des items de mathématiques s'est déroulée uniquement pendant la première semaine, tandis que la correction des items de sciences a eu lieu pendant la deuxième semaine. En tout, 108 correctrices et correcteurs y ont participé, à la fois les francophones et les anglophones.

Au total, environ 27 000 livrets d'évaluation ont été corrigés, dont environ 7000 en français et 20 000 en anglais.

Mise en liasse des livrets

Tous les livrets d'évaluation avaient été rassemblés en liasses avant la correction des items. Chaque liasse contenait 10 livrets d'évaluation provenant de diverses provinces. Lors de la séance de correction, les correctrices et correcteurs ont choisi une liasse et vérifié le code d'identification de l'élève de chaque livret. Les correctrices et correcteurs ne pouvaient obtenir aucune information sur l'identité des élèves à partir des livrets d'évaluation. (Puisque ceux-ci étaient identifiés à l'aide d'un code d'identification, l'évaluation préservait leur anonymat.) Par ailleurs, les correctrices et correcteurs ne pouvaient pas savoir de quelle province les livrets d'évaluation provenaient, afin d'éviter que la correction des items soit biaisée en favorisant une province au détriment d'une autre.

Feuilles de correction

Les feuillets détachables pour les questions à réponse construite se situaient à la fin des livrets d'évaluation. Tous les items à réponse construite étaient corrigés par des correctrices et correcteurs qui étaient eux-mêmes des éducatrices et éducateurs, parce que ces questions exigeaient un certain jugement personnel et s'appuyaient sur leurs connaissances dans la matière concernée. Les correctrices et correcteurs se sont appuyés sur les descriptions des guides de codage pour attribuer divers codes aux réponses des élèves et les ont enregistrés dans les feuilles de correction. À la fin de la séance de correction, les feuilles de correction ont été renvoyées au CMEC, où les données ont été numérisées et une base de données a été créée avec toutes les données de l'évaluation et des questionnaires.

Guide de la correctrice ou du correcteur

Avant la séance de correction, les correctrices et correcteurs se sont vu fournir le *Guide de la correctrice ou du correcteur* contenant des informations sur la logistique de la séance de correction et décrivant les responsabilités du personnel du CMEC, des chefs de correction et des correctrices et correcteurs. Ce guide fournissait également des informations sur le traitement des cas spéciaux, comme les points de vue tendancieux de la part de la correctrice ou du correcteur ou les suspicions de triche. Le *Formulaire de commentaires de la correctrice ou du correcteur*, qui devait être rempli à la fin de la séance de correction, était également inclus dans ce guide.

Guide de codage

Le *Guide de codage* fournissait une introduction générale sur l'art d'attribuer des codes aux réponses et des détails sur les principes du codage, par exemple des lignes directrices pour les fautes d'orthographe et de grammaire et des définitions des termes et des codes spéciaux. Le *Guide de codage* fournissait la catégorie de chaque question et une description de tous les codes possibles, ainsi qu'un éventail d'exemples de réponses susceptibles de recevoir un plein crédit ou un crédit partiel pour la question.

Chefs de correction

Les chefs de correction se sont rencontrés quelques jours en juin pour préparer la séance de correction. Ils ont examiné et adapté le matériel se rapportant à l'évaluation, notamment le *Guide de codage*. Ils ont également préparé le matériel de formation pour les correctrices et correcteurs. Lors de la préparation du matériel de formation, les chefs de correction ont choisi des échantillons de travaux d'élèves pouvant être utilisés comme exemples ou dans les documents de formation. Ces échantillons ont été numérisés et insérés dans le document de formation approprié. Certains échantillons choisis lors de la mise à l'essai ont également été inclus dans le matériel de formation. Ces échantillons servaient à illustrer la distinction entre les différents codes pour chaque item. Les chefs de correction avaient pour responsabilité d'assurer la formation des chefs de table et de garantir le bon déroulement de la séance de correction.

Chefs de table

Les chefs de table dirigeaient chacun une table de six à huit correctrices ou correcteurs. Ils avaient été formés par les chefs de correction. Leur rôle était de former les correctrices et correcteurs de leur table, de superviser leur travail, de renouveler la formation d'individus ou de groupes au besoin pour maintenir la cohérence du codage et de coder les copies.

Formation des correctrices et correcteurs

L'ensemble des correctrices et correcteurs, y compris les chefs de table, ont suivi une formation sur les guides de codage pour la lecture, les mathématiques ou les sciences, selon le rôle qui leur était attribué, avant de se lancer dans la correction des copies des élèves. Avant la séance de formation, les chefs de correction ont choisi des échantillons de travaux d'élèves à utiliser lors de la formation. Ils ont choisi ces exemples afin d'illustrer clairement les différences entre les codes possibles pour chaque question et ils ont procédé à un travail d'examen et de discussion sur ces exemples. Les copies de formation ont ensuite été utilisées par les correctrices et correcteurs pour s'exercer à la correction et pour mieux intérioriser le système de codage. Les correctrices et correcteurs se sont ensuite mis par deux pour corriger une liasse de livrets et ont discuté entre eux des codes retenus. Le processus a été répété pour plusieurs liasses, jusqu'à ce que les corrections soient conformes au guide. À la fin de la formation, lorsque les correctrices et correcteurs étaient capables d'appliquer les codes de façon cohérente, ils sont passés à un travail individuel de correction, jusqu'à ce que tous les livrets d'évaluation soient corrigés. Une fois que la correction d'un scénario était achevée, les correctrices et correcteurs ont reçu une formation sur la correction des items du scénario suivant.

Pour garantir un niveau élevé de cohérence dans la correction, une question a été codée dans l'ensemble des 10 livrets avant que les correctrices et correcteurs passent à la question suivante dans le bloc ou le scénario. Le processus a été répété jusqu'à ce que toutes les questions du bloc aient été codifiées. Lorsque les correctrices et correcteurs avaient fini de corriger une liasse, ils la rapportaient au chef de table et s'attaquaient à une autre liasse, jusqu'à ce que l'intégralité du bloc d'items, qui pouvait contenir un ou plusieurs scénarios, soit corrigée. Tout au long des processus de correction, les chefs de table ont procédé à une vérification aléatoire des codes attribués par chaque correctrice ou correcteur, afin de garantir la cohérence de l'adhérence aux guides de codage. Lorsque des problèmes se posaient pour des questions spécifiques, la solution était de procéder à une nouvelle formation soit individuelle soit collective et, dans quelques cas, à un nouveau codage de la question.

Les tables se sont vu attribuer soit des copies en anglais soit des copies en français. Les tables de correctrices et correcteurs bilingues, qui pouvaient aider soit l'équipe anglophone soit l'équipe francophone dans le travail de codage des items, ont été affectées soit aux copies en anglais soit aux copies en français, selon le secteur qui avait plus de livrets ou qui avançait plus lentement dans la correction.

Fidélité des corrections

Le but du processus sur la fidélité des corrections était de fournir des preuves objectives du niveau d'accord entre les différents membres de l'équipe de correction pour les items à réponse construite, afin de prouver que les corrections étaient uniformément fidèles aux guides de codage. Lors de la séance de correction, des données ont été rassemblées à partir d'examens de fidélité et pour la fiabilité entre membres de l'équipe de correction ou pour la double correction.

Examens de fidélité

Dans le cadre d'une séance de correction, il est toujours important d'instaurer les procédures nécessaires pour s'assurer que les correctrices et correcteurs codent comme il se doit. Il s'agit d'une étape majeure, car ils doivent tous s'entendre sur les différents codes afin d'assurer la validité des résultats. Avant la séance de correction, le personnel du CMEC a choisi des items au hasard dans l'ensemble des livrets d'évaluation reçus des provinces pour faire des examens de fidélité. Ces items choisis dans un ou plusieurs scénarios ont ensuite été distribués aux chefs de correction pour qu'ils en fassent le codage. Par la suite, leurs réponses ont été renvoyées au personnel du CMEC, qui les a saisies dans un fichier Excel servant à faire des comparaisons par rapport aux réponses des correctrices et correcteurs lors de la séance de correction. Si les chefs de correction mettaient en évidence un problème spécifique sur des questions particulières, des examens de fidélité supplémentaires étaient élaborés pour régler le problème. Les examens de fidélité ont donc fait office à la fois de contrôle de la qualité et de formation supplémentaire pour les correctrices ou correcteurs. Les examens de fidélité ont été exécutés pour l'ensemble des correctrices ou correcteurs francophones ou anglophones dans les trois domaines d'évaluation. Le but de ces examens était de surveiller la cohérence des corrections tout au long de la séance. Les examens de fidélité se sont produits environ deux fois par jour, selon la procédure suivante :

- À un moment fixé par le chef de correction, tout le monde s'arrêtait de corriger et corrigeait les mêmes échantillons de travaux d'élèves.
- Les codes attribués par les correctrices ou correcteurs étaient comparés aux données de référence (fournies par les chefs de correction).
- Les données étaient immédiatement saisies par le personnel du CMEC, qui fournissait les résultats au chef de correction.
- Les chefs de corrections faisaient le bilan avec l'ensemble du groupe ou avec des correctrices ou correcteurs à titre individuel.
- Si la cohérence était inférieure à 80 p. 100 pour une question spécifique, les individus ou groupes de correctrices et correcteurs suivaient une nouvelle formation et les livrets étaient corrigés une nouvelle fois, selon les besoins.

Les examens de fidélité ont donc permis de vérifier la concordance entre les résultats des experts et ceux des correctrices et correcteurs, c'est-à-dire qu'ils ont permis de vérifier si ces derniers avaient attribué les mêmes codes que les spécialistes pour les items d'un ou de plusieurs scénarios. Pour chaque examen de fidélité, il y avait un pourcentage du degré

d'accord calculé pour chaque correctrice et correcteur et pour chaque item. Il était attendu que le degré d'accord entre les spécialistes (chefs de correction) et les correctrices et correcteurs soit d'environ 85 p. 100. Si le résultat global des examens de fidélité était faible pour telle ou telle question ou tel ou tel bloc de questions spécifique, alors le groupe suivait une nouvelle formation et les copies déjà corrigées étaient vérifiées à nouveau par les chefs de correction, ou bien le codage de la question recommençait. Si le résultat des examens de fidélité était faible pour des correctrices ou correcteurs ou pour des tables spécifiques, alors les chefs de table offraient une nouvelle formation à l'individu ou au groupe concerné avant de passer à la correction. Les items précédemment corrigés par ces correctrices ou correcteurs étaient vérifiés.

Le pourcentage du degré d'accord par correctrice ou correcteur a été déterminé à l'aide du calcul suivant :

$$\text{Pourcentage du degré d'accord} = \frac{\text{Nombre total de réponses convenues}}{\text{Nombre total de tests de fidélité}} \times 100.$$

À la fin de la séance de correction, tous les pourcentages obtenus pour chaque examen de fidélité de chaque correctrice et correcteur ont été compilés, ce qui a constitué le degré d'accord total en pourcentage. Les tests de fidélité en lecture ont été répartis en deux groupes, puisqu'il s'agissait du domaine principal; il a donc fallu corriger un plus grand nombre de ces items. Les résultats nous montrent que la majorité des correctrices et correcteurs ont obtenu un niveau d'accord en pourcentage avec les spécialistes qui était plus qu'acceptable. Les résultats aux examens de fidélité ont été satisfaisants pour tous les groupes en mathématiques et en sciences. L'un des deux groupes en lecture a obtenu un niveau global d'accord en pourcentage inférieur à 80 p. 100. Les tableaux 6.1 à 6.4 suivants présentent le niveau d'accord en pourcentage par langue et globalement pour chaque groupe de correction et chaque domaine.

Tableau 6.1 – Résultats des examens de fidélité pour le groupe de correction en lecture n° 1

Test de fidélité	Bloc 1							Bloc 2						Niveau d'accord en %
	1	2	3	4	5	6	7	1	2	3	4	5	6	
Anglais	60	76	73	77	71	88	87	56	77	72	80	77	77	76
Français	74	79	88	69	86	84	-	80	78	88	69	81	90	71
Global	67	78	80	73	78	86	87	68	78	80	74	79	84	74

Tableau 6.2 – Résultats des examens de fidélité pour le groupe de correction en lecture n° 2

Test de fidélité	Bloc 1			Bloc 2			Bloc 3				Niveau d'accord en %
	1	2	3	1	2	3	1	2	3	4	
Anglais	89	84	86	94	94	94	97	97	95	89	92
Français	81	80	81	97	97	-	94	96	-	-	89
Global	85	82	83	96	96	94	96	97	95	89	91

Tableau 6.3 – Résultats des examens de fidélité pour le groupe de correction en mathématiques

Test de fidélité	Bloc 1	Bloc 2			Bloc 3		Bloc 4		Niveau d'accord en %
	1	1	2	3	1	2	1	2	
Anglais	99	97	94	98	91	97	90	99	96
Français	95	97	95	94	100	90	99	97	96
Global	97	97	95	96	96	93	94	98	96

Tableau 6.4 – Résultats des examens de fidélité pour le groupe de correction en sciences

Test de fidélité	Bloc 1			Bloc 2		Bloc 3		Bloc 4		Niveau d'accord en %
	1	2	3	1	2	1	2	1	2	
Anglais	83	95	90	88	83	66	72	82	79	81
Français	84	94	81	88	81	69	69	79	73	80
Global	84	94	85	88	82	68	70	81	76	81

Fiabilité entre membres du groupe de correction (double correction)

La double correction est un indicateur de contrôle de la qualité pour la correction des livrets d'évaluation en lecture, en mathématiques et en sciences; elle consiste à prendre environ 2870 livrets (environ la moitié des livrets dans chaque langue) et de les faire corriger une deuxième fois par une autre correctrice ou un autre correcteur. Le tableau 6.5 présente le niveau global de concordance entre membres du groupe de correction pour les trois domaines. En lecture, le niveau d'accord pour le groupe n° 1 était plus bas que prévu (66 p. 100), tandis que le niveau global d'accord pour le groupe n° 2 était acceptable (86 p. 100). Comme le niveau de fiabilité et le niveau de fiabilité entre correctrices et correcteurs ont été plus bas que prévus pour le groupe de correction en lecture n° 1, un élément d'ancrage contextuel en lecture sera retiré des futures éditions du PPCE.

Tableau 6.5 – Niveau global d'accord entre membres du groupe de correction pour la double correction

Livret	Domaine et groupe de correction			
	Groupe de correction en lecture n° 1	Groupe de correction en lecture n° 2	Mathématiques	Sciences
1	64	81	96	90
2	65	88	92	86
3	63	90	94	86
4	72	*	96	83
Moyenne	66	86	95	86

* En raison du plus petit nombre d'items corrigés en lecture pour le livret 4, tous les items ont été répartis dans le groupe de correction en lecture n° 1 pour cette analyse.

Fiabilité des tendances

La fiabilité des tendances est un indicateur de contrôle de la qualité servant à estimer le niveau d'accord entre les correctrices et correcteurs en lecture, en mathématiques et en sciences pour les items d'ancrage à l'évaluation du PPCE 2016 et pour les mêmes items à l'évaluation du PPCE 2013. Quatre items pour la lecture, huit items pour les mathématiques et six items pour les sciences étaient communs aux deux éditions. Environ 1910 livrets de 2013 (près de la moitié dans chaque langue) ont été corrigés une deuxième fois par une autre correctrice ou un autre correcteur. Lorsque les items étaient présents dans plus d'un livret, seul un livret a été corrigé pour la fiabilité des tendances. Les livrets ont été corrigés tel que déterminé par les chefs de correction. La correction pour la fiabilité des tendances s'est faite en même temps que le travail de correction principal, mais au début du processus, afin de pouvoir utiliser les données pour faire concorder les scores entre les deux évaluations si cela s'avérait nécessaire.

Rapports et commentaires

Divers rapports ont fourni des preuves des forces et des faiblesses du programme, ce qui pourrait servir à améliorer les futures éditions du PPCE. Les coordonnatrices et coordonnateurs des écoles ont produit des rapports sur le processus administratif. Ces informations ont été récapitulées et incluses dans les rapports sommaires des coordonnatrices et coordonnateurs des provinces. Les correctrices et correcteurs ont fait des commentaires lors de la séance de correction. Les informations recueillies dans ces rapports sont résumées dans les sections qui suivent.

Rapport de la coordonnatrice ou du coordonnateur de la province

À la suite du déroulement de l'évaluation, les coordonnatrices et coordonnateurs des provinces ont préparé un rapport synthétisant les détails relatifs à l'évaluation et les renseignements fournis par les coordonnatrices et coordonnateurs des écoles. Ce rapport visait à récapituler les commentaires des écoles sur le déroulement de l'évaluation. Les informations recueillies à partir des rapports des provinces ont été utilisées pour apporter les changements nécessaires au processus administratif des évaluations à l'avenir. Le *Rapport de la coordonnatrice ou du coordonnateur de la province* incluait sept questions.

Tout d'abord, les coordonnatrices et coordonnateurs des provinces devaient résumer les moyens utilisés par les écoles pour inciter les élèves à participer sérieusement à l'évaluation. Dans la plupart des cas, les provinces ont envoyé des informations sur l'évaluation du PPCE aux parents et tuteurs ou tuteurs des élèves sélectionnés pour les encourager à y participer. En outre, les coordonnatrices et coordonnateurs des écoles ont rencontré les élèves sélectionnés avant le test pour discuter avec eux du but et de l'importance de l'évaluation, afin qu'ils fassent de leur mieux. Ils leur ont aussi précisé que l'évaluation était anonyme et que leurs résultats ne seraient pas comptabilisés dans le calcul de leur note pour le bulletin scolaire. À la fin de l'évaluation, certains élèves ont été récompensés de leur participation; plusieurs écoles ont offert un petit-déjeuner gratuit, des collations, des coupons pour la cafétéria, etc. Certaines

écoles ont également offert aux élèves des incitations externes, comme des cadeaux, des certificats ou des privilèges spéciaux, ou bien les ont remerciés en organisant des événements.

Le *Guide d'information à l'intention de l'école* décrivait les procédures administratives pour les tests du PPCE. Malheureusement, les responsables du test ne se sont pas tous familiarisés avec ce guide à l'avance. Le guide encourageait, par exemple, le personnel enseignant à accorder aux élèves de petites pauses si nécessaire tout au long de l'évaluation, mais les enseignantes et enseignants ne semblaient pas tous être au courant de cette mesure d'adaptation. Les instructions indiquaient que les élèves avaient le droit d'utiliser une calculatrice, des articles à manipuler et un dictionnaire (qui pouvait être un dictionnaire anglais-français) ou un dictionnaire de synonymes. Malheureusement, encore une fois, il semble que ces outils n'aient pas été mis à la disposition des élèves dans toutes les écoles. Même si les provinces avaient convenu de faire participer les élèves d'immersion française au PPCE en anglais, de façon à ce que les élèves répondent à l'évaluation en lecture dans leur langue première, certaines écoles ont choisi de faire participer ces élèves à l'évaluation en français. Ce choix se justifiait par le fait que certaines écoles avaient indiqué que les élèves d'immersion ayant participé à l'évaluation en anglais n'étaient pas bien familiarisés avec le vocabulaire en anglais employé dans les sections de mathématiques et de sciences du test, puisque l'enseignement de ces matières en salle de classe s'était déroulé en français.

Les coordonnatrices et coordonnateurs des provinces devaient également résumer les problèmes rencontrés par les écoles lors du déroulement de l'évaluation. Un petit pourcentage de livrets avait une page qui manquait ou qui était en double ou encore des agrafes mal placées. Il y a eu une certaine confusion chez les élèves face aux questions leur demandant de choisir « oui » ou « non », puis de justifier leur choix, parce que certains élèves ont pensé qu'ils pouvaient soit faire leur choix soit le justifier. Les enseignantes et enseignants ont indiqué qu'ils avaient aidé leurs élèves à bien comprendre ces énoncés. Certaines écoles ont reçu soit des livrets soit des questionnaires dans la mauvaise langue. Les coordonnatrices et coordonnateurs des provinces sont parvenus à régler certains de ces problèmes quand ils se sont présentés, parce qu'ils avaient des livrets supplémentaires dans les deux langues.

Les commentaires des provinces semblent indiquer que la majorité des écoles se sont conformées aux procédures administratives. Dans la plupart des provinces, un pourcentage élevé d'écoles a indiqué que l'évaluation s'était déroulée de façon excellente ou satisfaisante. Les écoles qui étaient seulement plus ou moins satisfaites du déroulement de l'évaluation ont expliqué que cela était dû au fait qu'elles avaient reçu les documents administratifs en retard ou que le nombre d'exemplaires des livrets d'évaluation était incorrect. Certaines écoles ont aussi fait part de préoccupations concernant la spécificité et la clarté des informations relatives au déroulement de l'évaluation. Il y a eu des écoles qui n'étaient pas suffisamment équipées pour gérer une séance d'évaluation d'une durée de deux à trois heures et qui auraient eu besoin d'un plus grand soutien durant l'évaluation.

Il semble que l'attitude des élèves ayant participé à l'évaluation ait été généralement positive. Quant à ceux qui avaient une attitude plutôt négative, soit ils n'ont pas compris l'importance de l'évaluation, soit ils ont été déçus d'avoir dû manquer des activités (par exemple des activités sportives) pour pouvoir participer à l'évaluation.

En règle générale, les coordonnatrices et coordonnateurs des écoles ont été satisfaits du *Guide d'information à l'intention de l'école*. Ils ont mentionné que les informations et les directives étaient claires et précises et qu'elles avaient facilité le déroulement de l'évaluation. Quelques-uns ont fait part de suggestions pour améliorer le déroulement de l'évaluation du PPCE à l'avenir. Comme ils trouvaient le matériel trop volumineux et détaillé, certaines enseignantes et certains enseignants ont suggéré de résumer les points essentiels en une ou deux pages, en employant un langage très direct et une mise en page point par point. Bon nombre d'enseignantes et enseignants ont aussi souligné qu'il faudrait que le document précise clairement si l'utilisation d'une calculatrice est autorisée ou non et qu'il donne des instructions précises concernant les livrets d'évaluation supplémentaires.

Les coordonnatrices et coordonnateurs des provinces ont suggéré des modifications de l'évaluation pour les élèves ayant des besoins spéciaux ou pour les élèves ayant des problèmes en ce qui concerne la langue de l'évaluation. Selon leurs rapports, seules les modifications dont bénéficient habituellement les élèves ayant des besoins spéciaux ont été adoptées, dont les suivantes : mise à disposition de scribes, mise à disposition d'assistantes et assistants en éducation pour l'aide à la lecture, disponibilité d'un lieu adapté, octroi de temps supplémentaire, impression en gros caractères et mise à disposition de dictionnaires français-anglais pour les locutrices ou locuteurs non natifs.

Les commentaires des coordonnatrices et coordonnateurs dans le rapport ont été très positifs et il semble que le processus administratif pour l'évaluation se soit bien déroulé. Les suggestions et les commentaires ont été pris en compte, en vue d'améliorer le processus pour les évaluations à l'avenir.

Rapport de la coordonnatrice ou du coordonnateur de l'école

Après l'évaluation, les coordonnatrices et coordonnateurs des écoles devaient remplir un rapport pour indiquer si la séance d'évaluation s'était bien déroulée. Le *Rapport de la coordonnatrice ou du coordonnateur de l'école* incluait 13 questions. Les commentaires des coordonnatrices et coordonnateurs aideront à mieux planifier le déroulement des évaluations à l'avenir et ont aidé à recueillir des informations sur la manière dont l'évaluation du PPCE s'est déroulée dans les écoles.

Tout d'abord, les coordonnatrices et coordonnateurs des écoles devaient décrire les mesures prises pour inciter les élèves à prendre au sérieux leur participation à l'évaluation. Les mesures utilisées pour encourager les élèves à faire de leur mieux étaient semblables à celles décrites précédemment par les coordonnatrices et coordonnateurs des provinces. Plusieurs coordonnatrices et coordonnateurs des écoles ont mentionné qu'ils avaient passé du temps avec les élèves sélectionnés pour les sensibiliser à l'évaluation et leur expliquer l'importance de faire de leur mieux au test et de répondre aux questions de façon sérieuse. Certains d'entre eux ont même rencontré les élèves quelques jours avant l'évaluation et passé en revue les exemples de questions avec eux. D'autres leur ont expliqué que les données obtenues permettraient de comparer les résultats des différentes provinces et qu'il était important de bien représenter son école et sa province. Certains autres ont envoyé des informations sur l'évaluation du PPCE aux parents et tuteurs ou tutrices des élèves sélectionnés pour

l'évaluation, pour qu'ils sachent quel était le but de l'évaluation et qu'ils puissent eux-mêmes motiver leur enfant. Certains coordonnateurs et coordonnatrices des écoles ont mentionné que certains élèves avaient été récompensés à la fin de l'évaluation par une collation gratuite, comme une pizza, ce qui a semblé être une source de motivation les encourageant à faire de leur mieux au test.

Les coordonnatrices et coordonnateurs des écoles devaient préciser dans leur rapport s'ils avaient apporté des modifications aux modalités d'évaluation pour les élèves ayant des besoins spéciaux. Certains ont dû accorder du temps supplémentaire à ces élèves pour qu'ils terminent l'évaluation. D'autres élèves ont été placés dans une autre salle de classe ou dans un endroit plus tranquille pour qu'ils arrivent mieux à se concentrer. De plus, certains élèves se sont vu fournir une personne pour leur lire les questions (mot pour mot) ou un scribe.

Les coordonnatrices et coordonnateurs des écoles devaient aussi indiquer s'il y avait eu des problèmes pendant le déroulement de l'évaluation. La plupart d'entre eux ont mentionné que la séance d'évaluation s'était bien déroulée, même si, pour certains, divers problèmes sont survenus lors de l'évaluation. Avant l'évaluation, les coordonnatrices et coordonnateurs des écoles ont reçu la *Liste des élèves* dans le *Formulaire de suivi des élèves*, avec un numéro d'identification pour chacun. Certains surveillants et surveillantes d'examen n'ont pas lu attentivement les consignes et ont distribué les livrets d'évaluation de façon aléatoire et ne se sont rendu compte que plus tard que les livrets portaient des numéros les assignant à des élèves spécifiques. Il y a donc eu de la confusion pour savoir à qui appartenait chaque livret. Ils ont aussi remarqué que certains élèves étaient fatigués et n'étaient pas motivés pour faire le test, car l'évaluation avait lieu à la fin de l'année (la participation aux tests ayant eu lieu, dans plus de 80 p. 100 des cas, au mois de mai) et les élèves avaient plusieurs examens de fin d'année. Par conséquent, certains coordonnateurs et coordonnatrices des écoles ont exprimé leur crainte que certains élèves n'aient pas pris le test au sérieux. D'autres coordonnatrices et coordonnateurs des écoles ne savaient pas que les élèves pouvaient avoir accès à des ressources matérielles (par exemple, dictionnaire unilingue ou bilingue, articles à manipuler, dictionnaire de synonymes ou calculatrice) et ont recommandé de mieux le préciser avant le déroulement du test.

Les coordonnatrices et coordonnateurs des écoles devaient indiquer les procédures d'évaluation qu'ils n'avaient pas pu respecter. La plupart d'entre eux ont respecté les procédures, puisqu'aucun problème n'est survenu lors de l'évaluation; toutefois, certains d'entre eux n'ont pas été en mesure de suivre les procédures, comme l'indique la liste ci-dessous :

- Certaines écoles ont choisi de ne pas aviser les élèves ni les parents et tuteurs ou tutrices du test avant la date d'évaluation.
- Par inadvertance, certaines écoles n'ont pas tenu compte de la consigne relative aux numéros de code d'identification des livrets d'évaluation (soit la correspondance entre les codes d'identification spécifiques et les noms des élèves). Une fois les problèmes cernés, les erreurs ont néanmoins été corrigées.

- Certaines personnes chargées de la surveillance ont perdu les formulaires de suivi des élèves et distribué les livrets d'évaluation au hasard. Il y a donc eu de la confusion pour savoir à qui appartenait chaque livret.
- Certaines écoles n'ont pas rempli correctement les formulaires de suivi des élèves, ce qui a créé des problèmes par la suite.
- Du fait que plusieurs écoles ont estimé que la durée du test était trop longue par rapport au niveau de concentration de certains élèves ou en raison d'événements spéciaux qui avaient lieu à l'école, l'évaluation a été divisée en deux séances, au lieu de se dérouler en une seule séance de 90 minutes avec une petite pause.

Les coordonnatrices et coordonnateurs des écoles devaient commenter les scénarios d'introduction (par exemple, pertinence, niveau de difficulté, niveau d'intérêt) et les questions de l'évaluation (par exemple, formulation maladroite, plus d'une réponse correcte ou aucune réponse adéquate, question non adaptée à l'âge des élèves). La plupart des coordonnatrices et coordonnateurs des écoles ont précisé que le niveau des questions de l'évaluation était approprié et que les élèves se sont sentis motivés. Toutefois, plusieurs ont estimé que les questionnaires étaient trop longs et parfois répétitifs. Par ailleurs, un certain nombre de commentaires ont été faits à propos de la formulation des questions de l'évaluation.

Les coordonnatrices et coordonnateurs des écoles devaient également calculer le taux de participation. Si le taux de participation des élèves était inférieur à 85 p. 100, une séance de rattrapage devait être organisée. Selon les rapports des coordonnatrices et coordonnateurs des écoles, le taux de participation a été supérieur à 85 p. 100 dans presque toutes les écoles.

Formulaire de commentaires de la correctrice ou du correcteur

À la suite de la séance de correction des items de l'étude principale, environ 120 correctrices et correcteurs ont rempli un questionnaire destiné à recueillir leurs opinions et leurs commentaires, dans l'optique de faciliter la préparation des séances de correction et des évaluations à venir. Le *Formulaire de commentaires de la correctrice ou du correcteur* était divisé en trois sections : la première renfermait les renseignements personnels des correctrices et correcteurs; la deuxième était axée sur le processus de correction; et la troisième portait sur l'instrument d'évaluation.

En règle générale, les commentaires des correctrices et correcteurs étaient positifs sur les documents fournis, le local, le processus de correction et les documents d'évaluation.

Les correctrices et correcteurs étaient invités à examiner les questions d'évaluation afin de faire part de leurs observations, lors de la séance de correction, sur les forces et les faiblesses dans les trois domaines. Il leur était en outre demandé de dresser une liste d'idées fausses de la part des élèves dans leurs réponses. Les informations recueillies seront incluses dans un prochain numéro de *L'évaluation... ça compte!*, dans lequel deux passages de lecture de l'évaluation du PCE 2016 en lecture seront publiés avec des commentaires sur les travaux des élèves et des exemples de réponses.

Chapitre 7. Établir une norme de rendement

Chaque fois que le contenu ou les types d'items utilisés dans les tests subissent des modifications significatives, un travail d'établissement des normes est effectué. Lorsque l'évaluation ne change pas d'une édition à l'autre, les tests peuvent faire l'objet d'une égalisation psychométrique (c'est-à-dire de comparaisons et de rajustements statistiques), de façon à ce que les élèves soient confrontés à la même norme de rendement à chaque évaluation et soient traités de façon équitable. En 2016, la lecture a été le domaine principal du PPCE pour la deuxième fois, et des changements significatifs ont été apportés au cadre d'évaluation, de sorte qu'il a été nécessaire d'établir des normes de rendement.

Séances d'établissement des normes

Les séances d'établissement des normes ont eu lieu en février 2017 à Toronto. Les réunions ont été divisées en trois séances : une séance de formation des chefs d'une journée; deux journées de séances d'établissement des normes; et une séance de rédaction d'une journée, pour réviser les niveaux de compétence.

L'établissement des normes avait pour but de définir les niveaux de rendement pour l'évaluation en lecture du PPCE. Ces niveaux de rendement ont été décrits à l'aide de points de coupure répartissant le rendement des élèves selon différentes catégories. Le processus de normalisation était conçu de façon à produire ces points de coupure selon des calculs valables et systématiques, avec tout d'abord un groupe de spécialistes des matières, puis en faisant appel aux responsables de l'élaboration des politiques et aux autres parties intéressées, lors d'une phase d'examen. Deux points de coupure ont été définis pour distinguer trois niveaux de rendement. Le niveau 2 a été désigné comme le niveau de rendement acceptable pour les élèves de 8^e année/2^e secondaire.

Les participants ont fait les tests, les ont corrigés, ont examiné les descripteurs des niveaux de rendement (DNR) et se sont ensuite livrés à trois cycles d'examen des tests à l'aide de la méthode du signet (Cizek et Bunch, 2007). À l'issue des trois journées, les points de coupure recommandés par les spécialistes ont été envoyés aux coordonnatrices et coordonnateurs des provinces pour qu'ils les examinent. Nous décrivons ci-dessous la démarche suivie pour élaborer et prendre en note ces recommandations.

Sélection d'un groupe de spécialistes

Il était important pour le CMEC que toutes les provinces participent au processus et qu'elles aient l'occasion de prendre part à la définition des points de coupure. Chaque province a été invitée à désigner deux représentantes ou représentants ayant une certaine expertise en mesure et en évaluation, et en ce qui concerne le contenu du programme de lecture pour la 8^e année/2^e secondaire. Le comité d'établissement des normes se composait de 21 spécialistes. Le CMEC a recruté les spécialistes responsables de l'établissement des normes en s'appuyant sur les noms fournis par les coordonnatrices et coordonnateurs des provinces. L'un des critères les plus importants pour un tel comité est qu'il faut que ses membres présentent des caractéristiques démographiques pertinentes. C'est pour cela que le CMEC a mis sur pied un

comité à la composition équilibrée sur le plan du sexe, de l'expérience, de la langue et de l'emplacement géographique. Les 21 spécialistes comprenaient des enseignantes et enseignants travaillant auprès du groupe d'âge ciblé. Le CMEC a pris soin de faire en sorte que tant les locutrices et locuteurs francophones que les locutrices et locuteurs anglophones soient bien représentés.

Versions préliminaires des descripteurs des niveaux de rendement

Les descripteurs des niveaux de rendement (DNR) jouent un rôle important dans tout processus d'établissement des normes. Ces DNR décrivent ce que les élèves devraient savoir et être capables de faire à chacun des niveaux de rendement pour la 8^e année/2^e secondaire. Les DNR sont cruciaux dans le processus d'établissement des normes, parce qu'ils guident les spécialistes en les aidant à conceptualiser les différences de niveau de rendement entre élèves.

Les DNR du PPCE 2007, année pour laquelle la lecture a été le domaine principal pour la première fois, ainsi que les travaux de recherche sur les tests internationaux (par exemple, TEIMS, PISA) ont été examinés. Il s'agissait d'énoncés décrivant ce que les élèves des trois niveaux de rendement savaient et étaient capables de faire, et les spécialistes s'y sont référés tout au long du processus de normalisation, afin de se faire une bonne idée, dans leur travail, de ce que devait être le rendement de l'élève à chaque niveau. Les DNR ont été énoncés dans des termes correspondant au cadre d'évaluation en lecture du PPCE et, à cette première étape, les spécialistes ont fait des suggestions de révisions à apporter, en se consultant les uns les autres et sur les conseils de la personne responsable du dossier au CMEC. Lors de la réunion, la personne responsable a pris en note les révisions et les suggestions et les a projetées à l'écran, afin que tout le monde puisse bien les voir. Elle a ensuite intégré les formules suggérées pour produire des DNR révisés, sur lesquels les spécialistes se sont mis d'accord. À l'issue de ce processus, les DNR étaient prêts à être utilisés lors de la réunion de normalisation.

Sécurité des documents

Comme l'établissement des normes utilise des documents authentiques, il était crucial d'assurer un bon niveau de sécurité. À son arrivée à l'atelier, chaque spécialiste a reçu un code d'identification qui lui était propre. Tous les documents sécurisés contenaient les mêmes codes, pour que, lors de la distribution, l'exemplaire du document corresponde au numéro du spécialiste auquel il était distribué. Les spécialistes ont été informés qu'ils avaient pour responsabilité de s'assurer que les documents portant leur numéro restent confidentiels. Ils ont également été invités à signer un formulaire de confidentialité avant de pouvoir recevoir les documents sécurisés. Aucun document ne pouvait quitter les salles des séances organisées en parallèle à quelque moment que ce soit pendant la journée de travail.

Méthode du signet

La méthode du signet a été choisie par souci de continuité avec les séances antérieures d'établissement des normes du PPCE et ce, pour les raisons suivantes : cette méthode se prête à des évaluations dans des formats mixtes; elle permet aux personnes participantes d'examiner ensemble les items à réponse choisie et les items à réponse construite; et elle se fonde sur la

théorie des réponses aux items (TRI) et convient donc parfaitement aux approches d'évaluation axées sur cette théorie. Par rapport à d'autres méthodes d'établissement des normes, la méthode du signet nécessite moins de décisions et ces décisions sont par ailleurs plus simples. Pour ces raisons, la méthode du signet est une approche jugée efficace, pertinente et appropriée pour l'établissement des normes du PPCE.

Le format global de l'évaluation du PPCE 2016 combinait des items à réponse choisie (par exemple, des items à choix multiples [CM], des questions de type vrai ou faux et des questions de type oui ou non) à un nombre significatif d'items à réponse construite courte (RCSC) et à réponse construite longue (RCSL) ou de questions ouvertes. Les items RCSC étaient des items de lecture auxquels les élèves pouvaient répondre par une réponse brève qui était corrigée de façon dichotomique (codée 1 ou 0, selon qu'elle était correcte ou incorrecte). Les items à RCSL étaient des items comportant un, deux ou trois points et exigeant des élèves une réponse écrite plus longue. Le tableau 7.1 présente une ventilation du format des items.

Tableau 7.1 – Composition des livrets d'évaluation du PPCE pour les items de lecture

	Réponse choisie	Réponse construite courte (code 0 ou 1)	Réponse construite longue (code 0, 1 ou 2; ou code 0, 1, 2 ou 3)
Livret 1	24	1	5
Livret 2	23	3	2
Livret 3	18	3	5
Livret 4	15	1	2

Avec la méthode du signet, les spécialistes ont examiné les items du test dans un livret des items ordonnés (LIO), dans lequel tous les items des quatre livrets utilisés dans l'évaluation étaient triés par ordre de difficulté, le plus facile figurant à la première page et le plus difficile à la dernière. Les items à CM et à RCSC n'apparaissaient qu'une fois dans le LIO, mais les items à RCSL et les informations contextuelles apparaissaient une fois pour chaque point dans le score. Autrement dit, un item valant deux points apparaissait deux fois, la première fois avec un exemple de réponse valant un point, puis plus tard avec un exemple de réponse valant deux points. Un item valant trois points apparaissait trois fois dans le LIO. Chaque page contenait des informations essentielles sur l'item, comme sa position dans le LIO, sa position dans son livret d'origine, la difficulté de l'item et le point dans le score associé à l'item dans cette position. Le CMEC utilise généralement le modèle de la TRI à deux paramètres et le modèle généralisé du crédit partiel pour le calibrage des items et la construction du test. Les paramètres *b* (emplacement) des calibrages indiquaient les difficultés des items.

Pendant l'examen du LIO, il a été demandé aux spécialistes de déterminer l'item pour lequel les élèves de chaque niveau de compétence auraient deux chances sur trois d'obtenir une réponse correcte, et de marquer cet item dans le LIO à l'aide d'un signet, d'où la méthode du signet. Les items *avant* le signet reflètent le contenu du test que les élèves de ce niveau de compétence devraient maîtriser, tandis que les items *après* le signet devraient refléter le contenu du test qui

est difficile pour les élèves. Chaque fois que les spécialistes examinaient un item, ils étaient invités à réfléchir à la question suivante : « L'élève de ce niveau de compétence aurait-il deux chances sur trois de répondre correctement à cet item? » Si le spécialiste y répondait par l'affirmative, il passait à l'item suivant. S'il y répondait par la négative, il mettait en signet l'item en question. La probabilité de réponse 67 (PR67) a été calculée pour obtenir les points de coupure entre les niveaux de compétence 1 et 2, et entre les niveaux 2 et 3.

Procédure d'établissement des normes

Au total, 21 participantes et participants de l'ensemble des provinces et trois membres du personnel du CMEC ont pris part à la séance d'établissement des normes. Les participantes et participants ont été affectés à l'une des quatre tables prévues, soit deux tables francophones ou deux tables anglophones. Chaque table avait un chef et quatre ou cinq participantes ou participants. Le processus de définition des points de coupure a pris deux journées, avec une troisième journée réservée au peaufinage des descripteurs des niveaux de rendement.

Le groupe a suivi une présentation sur le PPCE, les procédures administratives, les caractéristiques des items et le cadre d'évaluation (ces informations étant tout particulièrement pertinentes pour les participantes et participants prenant part pour la première fois à un projet lié aux évaluations pancanadiennes du CMEC). La présentation portait également sur les points de coupure, la méthode du signet, les niveaux de rendement, l'emploi du temps pour la séance, les barèmes de correction des items à choix multiples et les guides de correction des items à réponse construite. Les participantes et participants n'avaient, pour la plupart, jamais utilisé la méthode du signet et il a donc fallu les informer du processus et de leurs tâches au fil des deux journées de la séance. Enfin, ils se sont vu fournir des informations sur les niveaux de rendement, pour mieux distinguer clairement les trois niveaux de rendement.

Les participantes et participants ont ensuite pris le reste de la matinée et une partie de l'après-midi pour se familiariser avec l'instrument d'évaluation et avec les documents pour la séance. Cette étape a pris du temps, mais était nécessaire pour que les membres du groupe examinent attentivement les documents afin de bien « maîtriser » l'évaluation. Les participantes et participants ont eu des discussions à leur table concernant les items de l'évaluation et la difficulté de ces items et ils ont eu l'occasion non seulement d'examiner les items, mais également d'y répondre et de corriger leurs réponses, de façon à se faire une meilleure idée des descripteurs des niveaux de rendement.

La première ronde de placement des signets a eu lieu avant la fin de la journée : les participantes et participants ont d'abord examiné indépendamment les uns des autres chaque item dans le LIO, puis ont discuté en groupe de leurs conclusions et des raisons pour lesquelles tel item était plus difficile que les items figurant à un rang inférieur dans le livret. Après cette discussion, chaque participante ou participant devait choisir un point de coupure – c'est-à-dire placer un signet dans le LIO à l'endroit précis où se trouvait la dernière question à laquelle un élève affichant un rendement au niveau précisé aurait deux chances sur trois de répondre correctement. Le CMEC a compilé les réponses de l'ensemble des participantes et participants en notant dans un fichier Excel les numéros de page du LIO indiquant les deux points de

coupure. C'est la moyenne de l'ensemble des réponses qui a défini l'emplacement du point de coupure entre le niveau 1 et le niveau 2, et entre le niveau 2 et le niveau 3.

La deuxième journée a commencé par une discussion en séance plénière sur la première ronde de placement des signets. Le personnel du CMEC a affiché les résultats avec un graphique illustrant tous les numéros de page du LIO notés par les spécialistes aux deux points de coupure et un tableau de statistiques à chaque point de coupure avec la difficulté moyenne, la moyenne des numéros de page et l'éventail des numéros de page. Il y avait de grandes variations évidentes entre les réponses des participantes et participants, surtout pour le point de coupure entre le niveau 1 et le niveau 2, certains situant le premier point de coupure au tout début du LIO et d'autres le situant bien plus loin dans le livret. Ces variations ont débouché sur d'importantes discussions pertinentes, les membres du groupe expliquant les uns aux autres pourquoi ils avaient placé le signet sur tel ou tel item particulier. Plusieurs des participantes et participants ont indiqué qu'ils avaient eu de la difficulté à placer le premier signet, parce qu'ils trouvaient que certains items étaient plus faciles pour les élèves alors que les données indiquaient le contraire. Ces items ont par conséquent été situés à un rang plus avancé dans le livret. Pour le deuxième point de coupure entre le niveau 2 et le niveau 3, il y avait un éventail plus restreint de numéros de page, indiquant un plus grand degré d'accord. La première ronde a été un bon exercice pour les spécialistes et leur a donné l'occasion d'échanger des commentaires et des opinions.

La deuxième ronde a été semblable à la première, avec les spécialistes plaçant des signets indépendamment les uns des autres dans le LIO pour déterminer les deux points de coupure et fournissant la justification de leur choix en groupe. L'objectif de la deuxième ronde est d'obtenir un ensemble de résultats qui soit plus cohérent que celui constitué lors de la première ronde. Le personnel du CMEC a compilé les résultats et en a fait part au groupe. Certaines participantes et certains participants ont décidé de changer la place de leurs signets, tandis que d'autres ont choisi de les laisser au même endroit. Il y a eu moins de variations dans les réponses pour les deux points de coupure que lors de la première ronde. Même si l'éventail du point de coupure moins élevé était plus restreint, des améliorations s'avéraient toujours nécessaires pour accroître la cohérence. Pour la deuxième ronde (contrairement à la première), les participantes et participants se sont vu présenter des données d'impact sur l'incidence de leurs choix, c'est-à-dire le pourcentage d'élèves se situant aux niveaux 1, 2 et 3. Selon les réponses fournies par le groupe lors de la deuxième ronde, 45 p. 100 environ des élèves se situaient au niveau 1, 48 p. 100 au niveau 2 et 7 p. 100 au niveau 3. Le fait pour les participantes et participants de voir ces données d'impact leur a permis de vérifier leurs choix par rapport aux résultats et de réajuster leurs points de coupure au besoin. Cela est important car les participantes et participants étaient des spécialistes en éducation qui avaient des connaissances pratiques au sujet des répartitions attendues du rendement des élèves. Les tableaux de la fréquence cumulée de la théorie des réponses aux items (TRI) pour la statistique thêta de chaque livret ont été dressés au préalable et utilisés lors des séances afin de déterminer la proportion d'élèves se situant en dessous de chaque point de coupure et la

proportion se situant au point de coupure et au-delà⁹. Cependant, les instructions pour le groupe étaient claires : les participantes et participants devaient placer les signets en fonction de la difficulté des items et non en fonction du pourcentage d'élèves qu'ils souhaitaient avoir à chaque niveau.

Lors de la troisième ronde, les participantes et participants ont placé des signets entre les niveaux dans le LIO pour la dernière fois, en gardant ou en modifiant leurs choix précédents. Les résultats de cette ronde ont été bien plus cohérents que ceux des deux rondes précédentes. D'après les réponses du groupe lors de cette troisième ronde, les pourcentages d'élèves se situant à chaque niveau de rendement se présentent sur le tableau 7.2.

Tableau 7.2 – Répartition des élèves par niveau de rendement en lecture

Niveau de rendement	Niveau 1	Niveau 2	Niveau 3
Score moyen en lecture	399 ou moins	400 à 602	603 ou plus
Pourcentage d'élèves	14	73	12

Un questionnaire a été distribué aux participantes et participants à la fin de la séance afin de rassembler des informations, des commentaires et des réactions sur le processus d'établissement des normes et sur la méthode utilisée, ainsi que sur l'instrument d'évaluation lui-même. La plupart des membres du groupe ont indiqué qu'ils avaient apprécié la séance, qu'ils avaient été à l'aise dans le processus, que la séance avait été une expérience enrichissante et que la méthode du signet était une manière juste et facile à comprendre de fixer les points de coupure. Les participantes et participants ont également indiqué, pour la majorité, qu'ils étaient satisfaits de l'organisation de la séance et des chefs et responsables de l'animation et ils ont formulé des commentaires favorables sur l'instrument d'évaluation. La plupart ont déclaré que les textes et les questions étaient appropriés et que l'évaluation en lecture était équitable pour des élèves de 8^e année/2^e secondaire.

Descripteurs des niveaux de rendement

Après le processus de normalisation, un sous-groupe du groupe de spécialistes a révisé les descripteurs des niveaux de rendement. Les membres de ce sous-groupe ont examiné tous les items se situant dans les intervalles définissant les trois niveaux de rendement. Ils ont élaboré, à partir de ces items, une description des connaissances et des habiletés caractérisant le rendement des élèves à chacun des trois niveaux¹⁰. Le niveau 2 est considéré comme le niveau acceptable ou « seuil de compétence », c'est-à-dire le stade à partir duquel les élèves commencent à montrer qu'ils possèdent le niveau de compétence en compréhension de l'écrit qui leur permet de faire face aux situations de la vie courante. Les élèves qui se situent au

⁹ Comme nous l'avons décrit plus haut, la statistique thêta a été ajustée pour la probabilité de réponse de deux sur trois.

¹⁰ Ces descriptions figurent dans le rapport public sur le PPCE 2016 (O'Grady, Fung, Servage et Khan, 2018), à l'adresse <https://www.cmec.ca/publications/lists/publications/attachments/381/pcap-2016-public-report-fr.pdf>.

niveau 1 présentent donc un rendement inférieur à celui attendu de la part des élèves de leur niveau scolaire.

Les niveaux de rendement sont ainsi présentés sous la forme du pourcentage des élèves atteignant chaque niveau. Les tâches se situant au bas de l'échelle (niveau 1) sont jugées plus faciles et moins complexes que les tâches au niveau le plus élevé (niveau 3) et cette progression relative à la difficulté et à la complexité des tâches s'applique à la fois à la lecture globalement et à chaque sous-domaine de l'évaluation.

Chapitre 8. Traitement des données du PPCE

Le traitement des données est un aspect important et assez complexe du projet, car il y a des étapes précises à suivre pour assurer la validité des résultats. Le CMEC a donc assemblé un comité consultatif technique – groupe de spécialistes en mesure et évaluation, ainsi qu’en statistiques – composé de membres reconnus dans leurs domaines respectifs à travers le Canada et dotés d’une expertise de grande ampleur en matière d’évaluations à grande échelle en éducation.

Collecte des données

Les livrets d’évaluation et les questionnaires ont été distribués lors de l’étude principale. Avant le nettoyage des données, au Canada, ce sont 27 484 élèves de 8^e année/2^e secondaire qui ont fait le test et répondu au *Questionnaire de l’élève*. Au total, 1470 enseignantes et enseignants d’anglais responsables des élèves participants ont répondu au *Questionnaire du personnel enseignant* et 1355 directrices et directeurs d’école ont répondu au *Questionnaire de l’école*. La collecte de données pour ces documents s’est déroulée sur une période de plusieurs semaines.

Saisie des données

Tout comme pour la mise à l’essai, lors de l’étude principale, les élèves ont dû noircir des cercles sur une feuille de réponses pour les items à réponse choisie ou rédiger leurs réponses en quelques phrases dans le livret d’évaluation pour les items à réponse construite. Une fois qu’ils avaient terminé l’évaluation, ils disposaient de 30 minutes pour répondre aux questions du *Questionnaire de l’élève*, à la fin du livret d’évaluation.

À l’issue du déroulement de l’étude principale, les provinces ont toutes envoyé les livrets d’évaluation, les feuilles de réponses et les questionnaires au CMEC pour la saisie des données. Le *Questionnaire de l’élève*, le *Questionnaire du personnel enseignant* et le *Questionnaire de l’école* contenaient des items à réponse choisie et n’avaient pas à être codés par des spécialistes. Ils ont donc été envoyés à une entreprise externe pour la saisie des données. Les livrets d’évaluation ont ensuite été expédiés à Ottawa (Ontario) pour la séance de correction, lors de laquelle les correctrices et correcteurs ont corrigé tous les items à réponse construite dans plus de 27 000 livrets. Les correctrices et correcteurs devaient attribuer un code à chaque item en noircissant les cercles appropriés sur une feuille de correction.

Deux techniques ont été utilisées pour la saisie des données. Les données des feuilles avec cercles noircis ont été saisies à l’aide d’un scanner. Pour les données des questionnaires, la technique utilisée a été la saisie manuelle.

Pour les données sur le rendement, les fichiers contenant des données illisibles ou les items avec de multiples réponses ont été analysés par un logiciel de reconnaissance optique des marques. Si, par exemple, les cercles d’une feuille de réponses donnée n’étaient pas suffisamment noircis, le programme indiquait qu’il s’agissait d’un fichier problématique. L’agente responsable des données a vérifié ces fichiers électroniques un par un et a fait la saisie des données manuellement.

Contrôle de la qualité des données saisies

Pour les données des questionnaires, toutes les données ont été saisies en double par l'entreprise chargée de la saisie des données. Les écarts, s'il y en avait, ont été signalés à un tiers, afin de se pencher sur les différences.

Nettoyage des données

Lorsque les données ont été reçues après la séance de correction, la première étape a été de vérifier la concordance entre la structure de la base de données et celle de la base de données du CMEC. L'agente responsable des données a mis en évidence toutes les variables, en ajoutant ou en supprimant des variables, au besoin. Des contrôles de cohérence ont été effectués pour les codes de participation, les données sur le rendement, les données des questionnaires ainsi que les données reçues de l'entreprise chargée de la saisie des données. Tous les écarts ont été vérifiés et confirmés. Les fichiers de données ont ensuite été utilisés aux fins des procédures de nettoyage et de recodage spécifiques.

Recodage général

Après que le centre de données du CMEC a mené l'enquête sur tous les écarts et introduit des corrections dans la base de données, les règles générales suivantes ont été appliquées aux incohérences non résolues dans la base de données du PPCE (il s'agissait généralement d'un très petit nombre de cas ou de variables par province, tout au plus) :

- Les incohérences non résolues concernant l'identification des élèves et des écoles ont conduit à la suppression du dossier dans la base de données.
- Les dossiers des élèves qui ne contenaient pas à la fois les données sur le rendement de l'élève et les données du questionnaire ont été corrigés à l'aide du code de participation approprié.
- Les dossiers doubles de données ont été repérés et un seul a été gardé, si les deux dossiers montraient une correspondance à 100 p. 100 de données identiques. Dans le cas des dossiers doubles pour lesquels il n'y avait pas de correspondance, des efforts ont été déployés pour consulter les livrets aux fins de clarification et de correction.
- Dans les très rares occasions où la source d'incohérences ne pouvait être cernée, les deux dossiers doubles ont été supprimés.

Examen des données d'échantillonnage

La dernière étape de nettoyage des données pour les données d'échantillonnage et de suivi s'est fondée sur l'analyse des fichiers de suivi (*Formulaire de suivi des élèves*, formulaire de suivi des livrets, etc.). Le CMEC a analysé les données d'échantillonnage et de suivi, les a vérifiées et, si nécessaire, a procédé à un recodage supplémentaire. Si, par exemple, une province avait un nombre plus élevé d'élèves dans une langue que ce qu'exigeait le cadre d'échantillonnage, alors les codes de langue pour les écoles ont été vérifiés et recodés selon les besoins.

Dernier examen des données et préparation de la base de données

Une fois que toutes les données avaient été saisies et examinées, les fichiers ont été compilés et fusionnés. Les bases de données ainsi obtenues ont ensuite été utilisées à des fins d'analyse préliminaire et de pondération. Pour les questionnaires, les rapports contenaient des statistiques descriptives sur chaque item du questionnaire. Pour les données sur le rendement des élèves, deux analyses – une analyse classique et une analyse du fonctionnement différentiel des items (FDI) – ont été réalisées. Ceci a permis de produire des informations sur les items du test qui semblaient avoir produit des résultats différents et sur toutes les données ambiguës qui subsistaient dans les questionnaires. Avec ces informations, on a corrigé la clé, au besoin, et procédé à un recodage supplémentaire des données ambiguës. Par exemple, si l'ambiguïté était liée à des erreurs d'impression ou de traduction, alors un code « sans objet » a été appliqué à l'item.

Le résultat du recodage (exigé à la suite de l'analyse initiale des données sur le rendement des élèves et des données des questionnaires) a été introduit dans les fichiers de données. Les pondérations des élèves, du personnel enseignant et des écoles ont été estimées par Statistique Canada simultanément, selon la répartition de la taille des échantillons et des ajustements pour la non-réponse. L'Annexe A donne de plus amples détails sur la pondération. Après pondération par Statistique Canada, les pondérations ont été envoyées au CMEC, et les pondérations finales ont été utilisées à des fins d'analyses plus approfondies et de liaison avec les évaluations.

Chapitre 9. Analyse des données sur le rendement des élèves

Ce chapitre décrit l'analyse des données sur le rendement des élèves du PPCE de 2016. Il décrit et fournit un calendrier détaillé de l'exécution et de la coordination des différentes tâches. Le plan d'analyse a compris les étapes suivantes :

1. analyse préliminaire
2. analyse des items :
 - i. analyse classique
 - ii. analyse de la TRI
 - iii. analyse du fonctionnement différentiel des items (FDI)
3. fonctionnement du test
4. liens et équivalences entre les résultats du PPCE de 2016 en lecture et en mathématiques et les résultats du PPCE de 2007, de 2010 et de 2013
5. correction et mise à l'échelle des données sur le rendement du PPCE de 2016
6. estimations des erreurs-types
7. présentation des résultats sur le rendement des élèves du PPCE de 2016

Analyse préliminaire

L'analyse préliminaire est venue s'inscrire en prolongement du processus de nettoyage des données. Elle a compris trois étapes : (1) filtrage des données; (2) recodage des items; (3) gestion des données manquantes. Ces étapes ont été effectuées pour chaque livret, en décomposant les informations selon la province et selon la langue. Cette décomposition a facilité le processus de vérification des données, par exemple, en mettant en évidence les cas intéressants d'items que l'élève n'avait pas atteints.

Filtrage des données

Des tableaux de fréquence ont été produits pour chaque item, en décomposant selon la province. Ces tableaux ont été utilisés :

- pour vérifier la présence d'anomalies dans les données (données aberrantes, clés incorrectes, etc.);
- pour examiner (examen de premier degré) la répartition des réponses;
- pour déterminer (et, au bout du compte, évaluer) le taux de données manquantes par item et par livret¹¹.

En outre, les données ont été nettoyées de façon à ce que les cas comportant des réponses vides soient supprimés. En raison de la conception des livrets du PPCE, les élèves ayant tenté de répondre à au moins un item à choix multiples (CM) et un item à réponse construite (RC) par matière ont été pris en compte à des fins d'analyses, pour s'assurer de disposer des données des élèves ayant fait l'effort de participer à l'évaluation. À ce stade du nettoyage, on a cerné les

¹¹ Les types de données manquantes et leur traitement sont décrits plus loin.

cas où les élèves ayant participé à l'évaluation en étaient exemptés (codes de participation 4, 5 ou 6) : il pouvait s'agir d'une exemption accordée par l'école, d'une exemption liée au fait que des modifications adéquates n'avaient pu être apportées ou d'une exemption découlant du fait qu'un élève donné ne fréquentait plus l'école. Par ailleurs, certains de ces élèves ont pu bénéficier d'adaptations. Une enquête plus approfondie a révélé que certains de ces élèves avaient suffisamment de données pour être retenus, ce qui veut dire qu'ils avaient tenté de répondre à au moins un item par type de question, par matière. Les cas d'identification erronée auraient pu être liés au fait que l'élève avait reçu un livret comportant un numéro d'identification qui ne correspondait pas à celui du *Formulaire de suivi des élèves*. Après avoir consulté les provinces, il a été décidé de maintenir ces élèves dans l'ensemble final de données. Cependant, les données correspondant à ces élèves n'ont pas été incluses dans le processus de calibrage.

La vérification des taux de participation pour toutes les provinces est effectuée durant le filtrage des données. Au cours de ce processus, il a été découvert que la proportion d'écoles privées par rapport aux écoles publiques au Québec ne reflétait pas la proportion de sa population. Le PPCE 2016 était la première édition pour laquelle les provinces étaient invitées à vérifier leurs taux de participation dès que les données avaient été reçues, et avant les analyses des données. La gouvernance scolaire n'étant pas une variable de stratification dans le cadre du PPCE, cet écart n'aurait pas été détecté avant que soient atteintes les étapes ultérieures des analyses intégrées aux précédentes éditions du PPCE. Statistique Canada a été mandaté pour effectuer une repondération des données afin de mieux refléter la population au Québec. Des renseignements détaillés sur le processus de repondération sont présentés à l'Annexe B.

Recodage des items

Avant de procéder aux analyses des données, les ensembles de données brutes du PPCE 2016 ont fait l'objet d'un recodage et d'un nettoyage, selon les analyses requises. Le recodage des données était requis pour le recensement des items valables, pour le recodage des différents types de réponses manquantes et pour les analyses de la TRI. Les données brutes de l'évaluation du PPCE 2016 comprenaient à la fois des réponses valables et des réponses non valables pour les items du test. Pour un item à choix multiples (CM), la réponse était valable si l'élève n'avait choisi qu'une option pour la réponse, que son choix soit correct ou non. La réponse était considérée comme non valable si l'élève avait choisi plus d'une réponse. Pour les questions ouvertes, la réponse construite (RCS) était traitée comme étant valable si elle était en rapport avec la question posée, indépendamment de la question de savoir si elle méritait un crédit plein ou partiel ou non. Si la réponse de l'élève était sans rapport avec la question, alors elle était considérée comme incorrecte. Des codes numériques ont été attribués aux items à CM et à RCS non valables.

Les items à CM des versions française et anglaise ont été recodés séparément. Cela s'est avéré nécessaire parce que les clés pour certains des items à CM en lecture, qui étaient des items d'ancrage repris des évaluations précédentes, n'étaient pas les mêmes dans les deux langues parce que les leurres étaient présentés hiérarchiquement sous la forme d'une pyramide. Si l'on

n'avait pas recodé ces items, cela aurait débouché sur des problèmes lors du processus de calibrage (par exemple, avec l'impossibilité d'atteindre la convergence).

Chaque option pour les réponses a été transformée en variable avec des valeurs binaires. Quatre nouvelles variables dichotomiques ont été calculées pour chaque item à CM. Ce nouvel ensemble de variables comprenait une variable pour la réponse correcte et une variable pour chacun des trois leurres. Ces variables ont été utilisées pour l'analyse classique des items.

Données manquantes

Comme pour les autres évaluations à grande échelle, trois types de données manquaient dans l'évaluation du PPCE 2016¹² :

- données manquantes en raison de l'échantillonnage de l'item (item non utilisé dans l'évaluation);
- réponse manquante parce que l'élève n'a pas eu le temps de terminer l'évaluation (item non atteint);
- items omis.

Pour distinguer ces trois types de données manquantes les uns des autres et des réponses multiples ou non valables, les codes suivants ont été utilisés :

- non utilisé dans l'évaluation : système manquant;
- sans objet : 7;
- non atteint : 6;
- omis : 9.

Items non utilisés dans l'évaluation

Les items *non utilisés dans l'évaluation* découlent de la conception de l'évaluation du PPCE, qui s'appuie sur la technique d'échantillonnage multi-matrices. Cette technique divise les items de l'évaluation en sections ou livrets, avec certains items communs à certaines ou à l'ensemble des sections. Chaque section est ensuite affectée à un sous-groupe distinct dans l'échantillon principal. Au PPCE, les questions ont été divisées en quatre livrets, avec certains blocs d'items communs à des paires de livrets. Comme chaque élève ne devait faire que certains des items du test, il n'y avait pas de réponse pour les items affectés aux trois autres livrets et les réponses étaient donc manquantes en raison de la conception même de l'évaluation. Par conséquent, les items non utilisés dans l'évaluation relevaient de la catégorie des données manquantes de manière complètement aléatoire (MCA). Ils peuvent, à ce titre, être ignorés et ont été traités comme des données manquantes.

¹² Le PISA a ajouté les réponses multiples ou non valables comme quatrième catégorie de données manquantes. Les réponses multiples n'ont pas été considérées comme des données manquantes au PPCE et ont été traitées comme des types différents de données.

Items sans objet

Le code « sans objet » a été utilisé si la question avait été mal imprimée, de sorte qu'il était impossible pour l'élève d'y répondre. Par exemple, il pouvait y avoir eu une erreur de photocopie ou d'impression rendant la question illisible. Le code « sans objet » n'a été utilisé que dans un petit nombre de cas et les items ont été traités comme des valeurs manquantes.

Items non atteints

Les items non atteints correspondent aux questions auxquelles l'élève n'a pas répondu et qui étaient regroupées vers la fin de l'évaluation. Cette situation se produit dans le vecteur de réponses de l'élève parce qu'il n'a pas eu le temps d'y répondre. Dans les évaluations internationales, un item est considéré comme non atteint quand l'élève n'y a pas répondu et n'a pas non plus répondu à l'item qui le précède immédiatement. Il faut en outre que l'élève n'ait tenté de répondre à aucun autre item dans le reste du livret¹³. Autrement dit, le premier item avec réponse manquante suivant le dernier item avec réponse valable (ou non valable) est traité comme celui que l'élève a tenté de faire, mais n'a pas eu le temps de terminer.

Pour le PPCE, les items non atteints ont été traités comme des items ignorés. Cette méthode est défendue par Lord (1980), pour qui il est impossible d'obtenir des informations faciles à quantifier à partir de tels items pour l'estimation des scores thêta (voir aussi de Ayala, 2009). Le PPCE 2016 a traité les items non atteints selon les approches utilisées par TEIMS et le PIRLS. Ces deux évaluations internationales les traitent comme des items non utilisés dans l'évaluation lors du calibrage des items. Dans l'estimation des scores thêta, cependant, ces items sont considérés comme des réponses incorrectes.

Items omis

Les items omis sont des items que l'élève a sautés lors de l'évaluation, soit par inadvertance soit parce qu'il ne connaissait pas la réponse. Ces items apparaissent plus tôt dans le test que les items non atteints, qui sont regroupés vers la fin. Lord suggère qu'il convient de ne pas ignorer les items omis (cité dans de Ayala, 2009, p. 150). Il soutient que la pratique consistant à ignorer les items omis, une estimation de niveau de compétence élevée risque d'être obtenue si l'élève ne répond qu'aux questions pour lesquelles il a confiance en sa capacité de répondre correctement. Même si le PPCE ne rend pas compte des scores individuels, les items omis ont reçu le même code que les réponses incorrectes.

¹³ Les items non atteints sont définis dans les rapports techniques du PISA, du PIRLS et de TEIMS. Pour le PIRLS et TEIMS, [*Traduction libre*] « un item est considéré comme non atteint lorsque [...] l'item lui-même et l'item qui le précède immédiatement sont sans réponse et aucun autre item n'est fait dans le reste [...] du livret » (Foy, Brossman et Galia, 2012, p. 18). Pour le PISA, les items non atteints sont [*Traduction libre*] « toutes les valeurs manquantes consécutives regroupées à la fin de la séance de test [...] sauf la première valeur de la série manquante, qui est codée comme étant manquante » (OCDE, 2012, p. 236 dans le rapport en anglais).

Réponse non valable

Il y a réponse non valable quand la personne interrogée choisit plus d'une réponse pour un item donné. Ce type de réponse a reçu le code 8.

Analyse des items

Deux familles d'analyses ont été exécutées : (1) analyse de la théorie classique des items et (2) analyse de la TRI.

Analyse de la théorie classique des items

L'objectif de l'analyse classique était de produire des statistiques pour un deuxième examen des items du PPCE 2016. Pour le domaine principal, c'est-à-dire la lecture, le premier examen s'est appuyé sur les données de la mise à l'essai. Les domaines secondaires consistaient en des items d'ancrage tirés des évaluations précédentes. Les items d'ancrage ont été utilisés pour les domaines secondaires afin d'évaluer les changements pour ces items au fil du temps (ou d'une cohorte à l'autre) en fonction de l'estimation de leur niveau de difficulté. Les items d'évaluation en mathématiques ont été utilisés en 2010 et en 2013, et les items d'évaluation en sciences ont été utilisés en 2013. Ces items ont fait l'objet d'une mise à l'essai lorsque les mathématiques et les sciences étaient le domaine principal de l'évaluation. Les statistiques ont été examinées en vue de la sélection des items à inclure dans le PPCE de 2016.

L'analyse de la théorie classique des items pour les items du domaine principal s'est concentrée sur les aspects suivants :

1. difficulté des items;
2. discrimination des items;
3. statistiques spécifiques pour les items à réponse choisie (RC) (items à choix multiple, items vrai ou faux, items oui ou non, etc.);
4. statistiques spécifiques pour les items à réponse construite (RCS);
5. pourcentage d'élèves choisissant chaque option de réponse pour chaque item;
6. pourcentage d'élèves n'atteignant pas l'item;
7. pourcentage d'élèves omettant l'item;
8. indices de fiabilité (c'est-à-dire indice de cohérence interne pour les items à RC et niveau d'accord entre membres de l'équipe de correction pour les items à RCS).

Ces statistiques ont été calculées pour chaque livret : quatre livrets pour la version anglaise du test et quatre livrets pour la version française. Il y a donc eu huit tests pour le domaine de la lecture. Il y a également eu huit tests, quoique de moindre ampleur, pour chaque domaine secondaire. Pour ces domaines secondaires, le placement des items dans l'évaluation du PPCE 2016 était conforme à leur placement dans l'évaluation initiale. L'effet de leur position a néanmoins lui aussi été évalué.

Difficulté des items

Pour chaque item à RC et pour les items à RCS dichotomiques, la difficulté correspondait à la valeur p classique. Pour les items à RCS polytomiques, c'est le pourcentage moyen qui indiquait leur difficulté. Dans les deux cas, les items non atteints ont été exclus du calcul.

Discrimination des items

Tant pour les items à RC que pour les items à RCS, la corrélation item-total corrigée – c'est-à-dire la relation entre la réponse correcte à un item et le score total – a été calculée. Le résultat attendu était une corrélation modérément positive entre les items avec de bonnes propriétés de mesure et l'échelle. Les items non atteints ont été exclus du calcul.

Statistiques spécifiques pour les items à CM

Pour les items à choix multiples, les statistiques spécifiques incluaient :

- le pourcentage d'élèves choisissant chaque leurre;
- la corrélation bisérielle de point entre chaque leurre et le score total pour tous les items donnés à l'élève dans un domaine donné. Pour les items ayant de bonnes propriétés de mesure, les leures présentaient une corrélation négative.

Statistiques spécifiques pour les items à RCS

Pour les items à réponse construite, les statistiques spécifiques incluaient :

- le pourcentage d'élèves dont la réponse se situait à chaque niveau;
- la corrélation bisérielle de point entre chaque niveau et le score total pour tous les items donnés à l'élève dans un domaine donné. Il était attendu que cette corrélation aille progressivement du négatif au positif avec l'augmentation des scores pour les items ayant de bonnes propriétés de mesure.

Examen pour mettre en évidence les données manquantes

Pour chaque item, les aspects suivants ont été examinés :

- le pourcentage d'élèves omettant l'item;
- le pourcentage d'élèves n'atteignant pas l'item;
- la corrélation bisérielle de point entre la variable omise de l'item et le score total pour tous les items donnés à l'élève dans un domaine donné;
- la corrélation bisérielle de point entre toute variable non atteinte de l'item et le score total pour tous les items donnés à l'élève dans un domaine donné.

Toutes ces statistiques ont également fait l'objet d'une estimation pour chaque population (ou province, lorsqu'un seul groupe linguistique était signalé) afin de pouvoir faire des comparaisons avec les estimations au niveau pancanadien.

Fiabilité de l'évaluation du PPCE 2016

Pour chaque domaine et sous-domaine, l'alpha de Cronbach a été utilisé comme indice de cohérence interne. Il a été calculé sur l'ensemble des livrets d'évaluation, en guise d'indice de fiabilité. Les moyennes de cet indice de fiabilité pour chaque domaine et sous-domaine ont également été calculées. La même chose a été faite pour chaque province.

Items problématiques

Les items problématiques ont été signalés à partir de l'analyse classique. Les items ont été signalés comme étant problématiques en présence de l'une ou plusieurs des conditions suivantes :

- corrélation bisérielle de point inférieure à 0,20;
- valeur p inférieure à 0,20;
- valeur p égale ou supérieure à 0,85;
- items plus faciles ou plus difficiles pour une province par rapport à la moyenne nationale¹⁴;
- corrélation bisérielle de point positive pour plus d'un leurre dans les items à CM ou corrélations bisérielles de point non ordonnées sur les différents niveaux pour les items à RCS;
- moins de 5 p. 100 des élèves sélectionnant un des leures dans les items à CM;
- moins de 10 p. 100 des élèves obtenant la valeur du score pour un item à RCS;
- accord entre membres du personnel de correction de moins de 70 p. 100 pour la valeur du score pour un item à RCS.

Dans le PPCE 2016, la classification d'un nouvel item à la fois dans les domaines de la lecture et des sciences a été effectuée par les spécialistes du contenu. C'est au moyen d'une analyse factorielle qu'il a été déterminé que cet item appartenait au domaine de la lecture. Les corrélations montrent également que le nouvel item est plus étroitement lié au score en lecture.

Analyse de la TRI

L'analyse de la TRI a fait intervenir les aspects suivants : (1) évaluation de la dimensionnalité du PPCE 2016; (2) estimation des paramètres des items; et (3) évaluation du caractère convenable des modèles de la TRI. L'évaluation des modèles de la TRI a fait intervenir la dépendance locale des items (DLI), la concordance entre la fonction mathématique du modèle et les données et l'invariance du PPCE 2016. Le processus s'est achevé par une évaluation du fonctionnement différentiel des items (FDI) du PPCE 2016 dans le cadre du traitement des données sur la validité.

¹⁴ Ceci suppose que le modèle de Rasch est adapté aux données afin de repérer les items et que l'analyse d'interaction des items par province est exécutée.

Évaluation de la dimensionnalité du PPCE 2016

La dimensionnalité du PPCE 2016 a été évaluée par analyse factorielle des items (AFI). L'AFI désigne la classe d'approches non linéaires pour déterminer la structure factorielle des données catégoriques (Cai, 2010). Ces approches sont plus appropriées que l'analyse factorielle classique, qui se fonde sur une matrice de corrélation linéaire entre les variables observées. Cette approche linéaire débouche sur l'extraction d'éventuels facteurs factices lorsqu'on traite des variables dichotomiques (ou polytomiques) (de Ayala, 2009; Laveault et Grégoire, 2002). Les approches non linéaires correspondent mieux, par conséquent, à ces types de données que les approches linéaires (McDonald, 1967).

Le programme de statistique IRTPRO a mis en œuvre une procédure du maximum de vraisemblance à informations complètes (MVIC) tenant compte de la non-linéarité entre les variables observées et entre les variables observées et la structure envisagée. Il a été conclu que l'hypothèse d'unidimensionnalité du domaine principal s'avérait fondée.

Calibrage des items et évaluation du caractère convenable des modèles de la TRI

Des items de paires de livrets ont fait l'objet d'un calibrage concomitant afin d'établir des liens entre tous les livrets et d'utiliser une échelle commune pour les scores. Cette procédure rend possible l'estimation des scores thêta sans que cela dépende de l'ensemble d'items auquel les élèves ont répondu. Les items des trois domaines ont fait l'objet d'un calibrage réalisé de façon indépendante, puisqu'ils mesuraient des matières différentes.

Trois modèles de la TRI ont été ajustés aux données simultanément. Pour les items à CM, la modélisation a ajusté le modèle à deux paramètres (2-PLM). Elle a ensuite été comparée au modèle à trois paramètres (3-PLM). Le modèle 2-PLM a été conservé parce que l'ajustement du modèle n'a pas produit d'amélioration significative lorsque le modèle 3-PLM a été testé; pour les items dichotomiques à CM, c'est le modèle 2-PLM qui a été utilisé. Les items polytomiques à RCS ont été calibrés à l'aide du modèle généralisé de crédit partiel (MGCP). Pour l'estimation de l'ensemble des trois paramètres des items, c'est la méthode d'estimation de la probabilité marginale maximale qui a été utilisée. L'évaluation du caractère convenable des modèles a fait intervenir la dépendance locale des items (DLI), à savoir la concordance entre la répartition des données empiriques et la répartition théorique (ou attendue).

L'évaluation de la DLI s'est faite au moyen de la statistique LD χ^2 (Chen et Thissen, 1997). Cette statistique est calculée en comparant les fréquences observées et attendues dans chacune des tabulations croisées bidirectionnelles entre les réponses à chaque item et les réponses à chacun des autres items. Ces statistiques de diagnostic sont (approximativement) des valeurs χ^2 normalisées qui augmentent en cas d'indication de dépendance locale entre deux items, c'est-à-dire si les données pour cette paire d'items indiquent une transgression de l'indépendance locale.

L'adéquation de la fonction mathématique spécifiée à la forme réelle des données a été évaluée à partir des statistiques $S-\chi^2$. (IRTPRO ne produit pas et n'approuve pas la production de la courbe de réponse aux items empirique.) Les statistiques $S-\chi^2$ se fondent sur la différence

entre les fréquences observées et les fréquences attendues dans les catégories de réponses selon les scores additionnés.

Fonctionnement différentiel des items

Pour le fonctionnement différentiel des items (FDI), il a fallu évaluer dans quelle mesure certains des items du PPCE 2016 présentaient des propriétés statistiques différentes (niveau de difficulté, etc.) selon le sexe et la langue. Autrement dit, le but du FDI est de déterminer si un item présente un biais par rapport à un sexe ou à un groupe linguistique. Cela s'est fait à l'aide de la méthode Mantel-Haenszel (M-H) et du test de Wald dans IRTPRO. Ce test se fait dans IRTPRO [*Traduction libre*] « avec des matrices exactes de variance-covariance des erreurs dans les paramètres des items calculées au moyen d'un algorithme espérance-maximisation complété » (voir la documentation technique du IRTPRO). Si certains des items ont bel et bien manifesté un fonctionnement différentiel, cela s'est fait de façon équilibrée entre les groupes comparés, comme l'a montré le chevauchement quasi parfait du fonctionnement différentiel du test.

Fonctionnement du test

Le fonctionnement du test a été évalué à partir du score moyen au test, de la variabilité des scores au test, d'un indicateur de cohérence interne (l'alpha de Cronbach), de l'erreur-type de mesure et de la fonction d'information du test.

Liens et équivalences avec les évaluations précédentes pour les domaines secondaires

Le travail d'établissement des liens et d'équivalences a donné une indication de l'évolution entre les évaluations précédentes et la dernière évaluation. L'ensemble des trois domaines du PPCE 2016 comportaient des items qui avaient été utilisés dans les évaluations précédentes, où ces domaines étaient le domaine principal. Aucun nouvel item n'a été élaboré dans les domaines secondaires pour le PPCE 2016. Par conséquent, tous les items en mathématiques et en sciences étaient des items d'ancrage. La conception de l'évaluation correspondait à la méthode méthode des groupes non équivalents avec ancrage. En raison du changement de définition de la population cible, l'année 2010 a servi d'année de référence pour l'ensemble du travail d'établissement de liens avec les cycles ultérieurs¹⁵. L'établissement de liens avec le PPCE 2016 s'est fait au moyen d'un calibrage concomitant avec le PPCE 2013. Comme les paramètres de deux items d'évaluation successifs étaient estimés simultanément avec des items d'ancrage communs au fil des ans, les paramètres des items d'ancrage avaient les mêmes estimations et se situaient sur la même échelle (de Ayala, 2009; Kim et Kolen, 2006). Cette

¹⁵ En 2010, la comparaison entre le niveau en lecture en 2007 et en 2010 s'est faite en utilisant comme valeurs de référence les paramètres des items de 2010 (voir CMEC, 2011). La décision d'utiliser 2010 comme année de référence au lieu de 2007 a été prise en raison du changement de population cible (les élèves de 8^e année/2^e secondaire au lieu des élèves âgés de 13 ans). Comme l'année de référence est devenue 2010, et pour garantir la cohérence du processus de comparaison, le calibrage a par conséquent utilisé l'échantillon de données de 2010 comme indicateur pour les tendances en lecture.

approche avait l'avantage d'effectuer une utilisation maximale de toutes les données disponibles pour l'estimation des paramètres des items (Martin, Mullis, Foy, Brossman et Stanco, 2012).

Pour ce qui est de l'estimation des scores thêta, des élèves des deux échantillons ont été utilisés pour définir l'échelle. Par conséquent, les scores obtenus par les élèves ayant participé à la dernière évaluation, quand ils faisaient l'objet d'une estimation à l'aide des paramètres des items obtenus dans le cadre du calibrage concomitant, ont été égalisés (de Ayala, 2009). Cependant, le recalibrage des items communs a signifié que leurs paramètres pouvaient changer au fil du temps. Comme les paramètres pouvaient changer au fil du temps, le PPCE 2016 a suivi d'autres programmes d'évaluation à grande échelle, comme le PIRLS et TEIMS, qui vont un cran plus haut pour incorporer cette évolution dans le processus d'établissement des liens. Plus précisément, l'approche du PIRLS et de TEIMS exige, une fois que le calibrage concomitant est fait, les étapes suivantes :

- estimer les distributions des niveaux de rendement pour l'évaluation actuelle en utilisant le paramètre du calibrage concomitant;
- déterminer la transformation linéaire qui fait le mieux concorder les distributions des niveaux de rendement de l'évaluation précédente estimées dans le cadre du calibrage concomitant aux mêmes distributions de l'évaluation obtenues lors de l'estimation des paramètres des items au cours du cycle précédent;
- appliquer la transformation linéaire de l'étape II aux distributions des niveaux de rendement de l'évaluation actuelle.

La production des scores pour le niveau de rendement en mathématiques et en sciences pour le PPCE 2016 (scores thêta et scores à l'échelle) a utilisé les paramètres des items estimés à ce stade.

Production des scores pour le rendement et scores à l'échelle

Pour chaque élève et dans chacun des trois domaines, la production des scores s'est faite en trois étapes :

- Un score thêta a été produit pour représenter le niveau global de l'élève dans le domaine en question (lecture, mathématiques ou sciences) à l'aide des paramètres des items obtenus à partir des calibrages concomitants. L'estimation du score thêta s'est faite à l'aide de la méthode « attente *a posteriori* ».
- Les scores à l'étape II ont été pondérés avec la pondération d'échantillonnage à l'échelle avec une moyenne canadienne de 500 et un écart-type de 100.
- Pour les sous-domaines de la lecture, tous les scores pondérés ont été rapportés à une base moyenne pancanadienne de 500 et à un écart-type de 100. Cela est dû à la séparation des liens entre 2007 (lorsque la lecture avait été le domaine principal) et 2010 (lorsque l'échantillon cible avait changé); les scores en lecture pour les sous-domaines n'ont pas été rééchelonnés en 2010. Les scores en lecture pour les sous-domaines du PPCE 2016 sont ainsi devenus la nouvelle base de référence pour les futures comparaisons des moyennes de rendement des élèves.

Estimations des erreurs-types

L'analyse des données du PPCE 2016 s'est servie d'une approche bootstrap pour produire de façon empirique les estimations des erreurs-types pour les résultats pancanadiens et les moyennes par province pour chacun des trois domaines évalués. Même si l'approche bootstrap est de plus en plus largement utilisée, en particulier dans les domaines de la recherche, elle peut être affectée par la production d'estimations qui ne sont pas cohérentes si la valeur de départ et le tri de variables changent à chaque exécution. Cela est dû au fait qu'il existe de nombreux échantillons aléatoires qui peuvent être tirés de l'échantillon initial. Par conséquent, si quelqu'un souhaite reproduire l'erreur-type pour les résultats du PPCE, il existe une certaine probabilité qu'il obtienne des résultats différents. Pour fournir aux lectrices et aux lecteurs une estimation plus précise des erreurs-types présentées dans les rapports du PPCE 2016, les valeurs de départ bootstrap et les variables de tri sont indiquées à l'Annexe C. Lors des estimations bootstrap, les pondérations des élèves ont servi à l'analyse de leur rendement.

Présentation des résultats de l'évaluation du PPCE 2016

Des rapports sommaires sur les scores ont été élaborés au niveau pancanadien, au niveau des provinces, selon la langue et selon le sexe pour chacun des trois domaines évalués. Les résultats ont été fournis sous forme de tableaux et de graphiques selon la présentation utilisée pour les rapports publics du PPCE de 2007, de 2010 et de 2013. Les intervalles de confiance de 95 p. 100 ont été calculés au moyen des erreurs-types bootstrap. Des tests T ont été réalisés pour toutes les comparaisons effectuées à l'aide des ajustements de Bonferroni, selon le nombre de comparaisons.

Chapitre 10. Analyse des données du questionnaire

Comme lors des évaluations précédentes, le PPCE 2016 a recueilli des informations contextuelles sur les élèves, le personnel enseignant et les écoles. Le *Questionnaire de l'école* a été rempli par les directions d'école. Le *Questionnaire de l'élève* a été rempli sous forme imprimée, avec les livrets d'évaluation. En ce qui concerne le *Questionnaire du personnel enseignant* et le *Questionnaire de l'école*, le personnel enseignant et les directions d'école ont eu la possibilité de remplir leur questionnaire sous forme électronique ou imprimée. L'analyse des données des questionnaires a compris les étapes suivantes :

1. analyse préliminaire;
2. statistiques descriptives;
3. analyse factorielle pour créer des variables calculées lorsque cela était approprié;
4. analyse des items pour les échelles élaborées postulées et empiriques;
5. analyse de comparaison des groupes;
6. analyse corrélacionnelle :
 - i. corrélation simple,
 - ii. modèles de régression linéaire multiple,
 - iii. modèles de régression à niveaux multiples.

Ces analyses statistiques ont été faites pour chacun des trois questionnaires et ont été présentées pour chaque langue. Le PPCE 2016 est le premier cycle dans lequel les pondérations du personnel enseignant ont été incluses; lors des précédents cycles, les pondérations des écoles avaient servi aux analyses au niveau du personnel enseignant, en partant du principe qu'il y avait une enseignante ou un enseignant par classe sélectionnée dans chaque école. Toutefois, cela n'a pas été le cas dans certaines écoles, ce qui a créé une certaine incohérence par rapport au nombre de membres du personnel enseignant et d'écoles participants.

Analyse préliminaire

L'analyse préliminaire a suivi la même procédure que pour les items de l'évaluation. Elle a compris un filtrage des données et le recodage de certains items. Le traitement des données non valables et des valeurs manquantes, cependant, a été légèrement différent. Il fallait s'attendre à des réponses non valables (c'est-à-dire les cas où l'individu donnait de multiples réponses à une seule et même question), des items omis et des items non atteints dans les données des questionnaires. Ces réponses ont toutes été traitées comme des valeurs manquantes. Cependant, les données n'étaient pas censées contenir d'items non utilisés dans l'évaluation, car tous les élèves avaient eu à remplir le questionnaire contextuel au complet.

Filtrage des données

Le filtrage des données a révélé certains cas où certains membres du personnel enseignant et certaines directions d'école avaient rempli le questionnaire à la fois sous forme électronique et imprimée, et leurs réponses comportaient des incohérences. Pour ces cas

comptés en double, des efforts ont été déployés pour déterminer la source d'erreurs. En ce qui concerne les personnes interrogées pour lesquelles il n'a pas été possible de déterminer cette source d'erreurs, c'est le cas qui comportait davantage de réponses qui a été conservé dans les ensembles de données. Dans une province, le volume des données associées au *Questionnaire du personnel enseignant* et au *Questionnaire de l'école* était moins élevé que prévu. Une enquête a révélé que les données découlant des questionnaires imprimés n'avaient pas été recueillies : une repondération du personnel enseignant et des écoles a donc été nécessaire pour cette province.

Des tableaux de fréquence ont été produits pour le filtrage de chaque item :

- pour vérifier la présence d'anomalies dans les données (par exemple, données aberrantes, erreurs);
- pour examiner la répartition des réponses (fréquence et pourcentage);
- pour déterminer le taux de données manquantes par item et par livret.

Recodage des items

Le PPCE 2016 comprenait à la fois des réponses valables et des réponses non valables. La réponse à une question était valable si l'élève n'avait choisi qu'une option pour la réponse, mais elle était considérée comme non valable si l'élève avait choisi plus d'une réponse. La tâche décrite ici a consisté à recoder les données brutes sur les réponses valables et non valables aux items des questionnaires de l'élève, du personnel enseignant et de l'école.

Les réponses non valables ont reçu le code 7 pour les distinguer des réponses valables et des réponses manquantes.

Certains items des questionnaires comportaient des réponses écrites, lesquelles devaient faire l'objet d'un codage. Des variables numériques étaient requises pour les analyses quantitatives de ces items. Ces items recodés incluaient les pays dans lesquels étaient nés les élèves et leurs parents, ainsi que les types de documents qu'aiment lire les élèves.

Données manquantes

Il y a eu trois types de données manquantes dans les questionnaires du PPCE 2016 :

- réponses manquantes parce que l'élève n'avait pas eu assez de temps pour remplir le questionnaire (item non atteint)¹⁶;
- items omis, c'est-à-dire items sautés par l'élève volontairement ou non à des endroits quelconques dans le questionnaire;
- données manquantes parce que les questionnaires des enseignantes et enseignants ou des élèves remplis sur papier sont retournés après la capture des données.

Ces types de données manquantes ont reçu le code 9. Quand cela était possible, les données

¹⁶ Il n'y avait pas de limite de temps pour le personnel enseignant et pour les directions d'école, de sorte que les résultats n'étaient pas censés avoir de données manquantes.

manquantes ont été saisies à l'aide du processus d'imputation multiple (IM). Les données manquantes présentent des problèmes significatifs dans la modélisation statistique, parce que le cas est typiquement supprimé s'il y a des données manquantes dans l'une quelconque des variables du modèle. Même si le nombre de cas manquants pour une variable donnée est faible, le nombre de cas manquants augmente de façon significative si les données manquantes sont éparpillées entre différents cas. L'utilisation d'une technique comme l'IM permet de minimiser le problème.

Statistiques descriptives

Les statistiques descriptives ont été produites par province et par langue. Elles comprenaient des distributions de fréquence et des distributions en pourcentage pour tous les items utilisant des échelles catégorielles ou de type Likert. Les statistiques descriptives incluaient également la moyenne, l'écart-type et les statistiques de forme (asymétrie).

Analyse factorielle

Dans les questionnaires, plusieurs périodes ont servi à l'examen de caractéristiques ou d'attitudes précises. L'analyse factorielle a fait intervenir l'exécution d'une analyse factorielle exploratoire (AFE) des items des questionnaires du PPCE 2016 en répartissant les groupes d'items dans des *facteurs*. Certains items ont été recodés afin de correspondre à des items semblables pour les analyses factorielles. Les scores factoriels ainsi obtenus ont été échelonnés sur la moyenne de 50 et l'écart-type de 10 pour devenir des *scores d'indice*. Ces scores d'indice ont ensuite été mis en corrélation avec le rendement en lecture et, en raison de la grande taille de l'échantillon, il a été établi que toutes les corrélations étaient importantes, même si, dans certains cas, le degré de corrélation était faible. Comme ce fut le cas pour les précédentes évaluations du PPCE, seuls les facteurs ayant des corrélations de .2 ou plus ont été présentés dans le rapport contextuel du PPCE 2016. Les scores d'indice ont également été divisés en quatre quartiles quasiment égaux comme suit :

- quartile inférieur : en dessous du 25^e centile
- troisième quartile : du 25^e au 49^e centile
- deuxième quartile : du 50^e au 74^e centile
- quartile supérieur : 75^e centile ou au-dessus

Analyses des items et des indices

Des analyses statistiques ont été réalisées pour les items et les indices des questionnaires du PPCE 2016. Une analyse a été effectuée pour chaque questionnaire et les résultats ont été présentés par province et par langue. Elle s'est concentrée principalement sur les éléments suivants :

- moyenne et écart-type;
- corrélations entre items;
- pourcentage de personnes interrogées ayant des réponses manquantes pour les items;

- corrélations et régressions entre scores d'indice et rendements des élèves;
- analyse de régression;
- alpha de Cronbach pour chaque indice.

Analyse de comparaison des groupes

L'analyse de comparaison des groupes a fait intervenir les éléments suivants :

- comparer les moyennes de rendement des élèves par rapport aux variables démographiques se rapportant à l'élève, au personnel enseignant et à l'école;
- comparer les moyennes de rendement des élèves par rapport aux indices se rapportant à l'élève, au personnel enseignant et à l'école;
- comparer les moyennes de rendement des élèves par rapport aux indices se rapportant à l'élève et au personnel enseignant par quartile;
- comparer les scores d'indice des élèves par sexe et par langue;
- comparer à l'échelle provinciale les moyennes des élèves par rapport aux moyennes pancanadiennes par indice se rapportant à l'élève;
- se baser sur les moyennes de rendement aux niveaux du personnel enseignant (c'est-à-dire la classe) et de l'école pour effectuer des comparaisons par rapport aux variables se rapportant au personnel enseignant et à l'école.

Analyse corrélacionnelle

L'analyse corrélacionnelle a fait intervenir les éléments suivants :

- calculer des coefficients de corrélation simple – aussi appelée corrélation bivariée ou corrélation d'ordre zéro – entre le niveau de rendement des élèves et une variable contextuelle ou une variable indice;
- effectuer une analyse de régression linéaire multiple pour prédire le rendement en lecture à partir d'un ensemble de variables se rapportant aux élèves;
- effectuer une analyse de régression linéaire multiple pour prédire le rendement au niveau de la classe, c'est-à-dire le rendement moyen de la classe en lecture, à partir d'un ensemble de variables se rapportant au personnel enseignant et à l'école;
- effectuer une analyse de régression linéaire multiple pour prédire le rendement au niveau de l'école, c'est-à-dire le rendement moyen de l'école en lecture, à partir d'un ensemble de variables se rapportant au personnel enseignant et à l'école.

Pour l'ensemble du travail d'analyse corrélacionnelle, on a supposé que la variable dépendante (le rendement de l'élève) était liée de façon linéaire aux variables explicatives. Cependant, les hypothèses de régression linéaire ont été vérifiées avant de procéder à l'analyse.

Chapitre 11. Ensembles de données du PPCE

Description des ensembles de données

Tous les ensembles de données du PPCE 2016 sont en français et en anglais, et sont à la disposition des chercheuses et chercheurs. Le CMEC a plusieurs ensembles de données pour le PPCE, dont l'un couvrant l'ensemble des élèves participants, l'un couvrant l'ensemble des écoles participantes et l'un couvrant l'ensemble des enseignantes et enseignants des élèves participants. Il y existe également un ensemble de données élèves/personnel enseignant/écoles qui contient tous les dossiers des élèves fusionnés aux réponses aux questionnaires. Cet ensemble de données permet d'établir des rapports entre le rendement des élèves et les données contextuelles. Les ensembles de données se présentent sous formats SPSS et Excel. Comme les étiquettes des variables ont été préparées en format SPSS, aucune liste de codage n'est fournie pour ce format; cependant, une liste de codage est fournie pour l'ensemble de données en format Excel.

Ensemble de données des élèves

Cet ensemble de données contient principalement les informations suivantes :

- renseignements généraux sur les élèves (numéros d'identification de l'élève, de l'école et de l'enseignante/enseignant; code de participation de l'élève; utilisation par l'élève des adaptations; numéro du livret d'évaluation; province et langue de l'élève);
- pondération statistique pour les élèves;
- réponses aux items du questionnaire de l'élève;
- scores d'indice associés aux élèves;
- scores de rendement et niveaux de rendement pour tous les domaines et pour les sous-domaines de la lecture.

Cet ensemble de données inclut tous les élèves ayant des scores de rendement. Il se peut néanmoins que certains cas ne comportent pas les réponses au questionnaire du fait du nombre moins élevé d'élèves ayant rempli le questionnaire, comparativement au nombre d'élèves ayant rempli les livrets d'évaluation.

Ensemble de données du personnel enseignant

Cet ensemble de données contient principalement les informations suivantes :

- renseignements généraux sur le personnel enseignant (numéros d'identification de l'enseignante/enseignant et de l'école; province et langue pour chaque enseignante/enseignant);
- pondération statistique pour le personnel enseignant;
- réponses aux items du questionnaire du personnel enseignant;
- scores d'indice associés au personnel enseignant¹⁷.

¹⁷ Les scores d'indice associés au personnel enseignant n'ont pas été inclus dans le rapport contextuel.

Comme des classes intégrales ont été sélectionnées, une enseignante ou un enseignant seulement a fait l'objet d'un échantillonnage dans la plupart des écoles, avec deux ou plusieurs enseignantes ou enseignants dans un petit nombre d'écoles.

Ensemble de données des écoles

Cet ensemble de données contient principalement les informations suivantes :

- renseignements généraux sur les écoles (numéro d'identification de l'école, province et langue de l'école);
- pondération statistique pour les écoles;
- réponses des directions d'école aux items du questionnaire de l'école;
- scores d'indice associés aux écoles¹⁸.

Ensemble de données fusionné – élèves/personnel enseignant/écoles

Cet ensemble de données contient toutes les informations des ensembles de données décrits ci-dessus. Il permettra aux chercheuses et chercheurs d'établir des rapports entre le rendement des élèves et les données contextuelles. Le nombre de cas figurant dans cet ensemble de données équivalra au nombre de cas inclus dans l'ensemble de données des élèves. Il convient de noter que, selon le type d'analyse, l'ensemble de données fusionné, établi au niveau de l'élève, peut ne pas être approprié pour toutes les analyses. Les bonnes pondérations devraient aussi être utilisées selon que le niveau de variables d'intérêt est établi au niveau de l'élève, du personnel enseignant ou de l'école.

Consultation de l'ensemble de données pour les recherches

Le PPCE, qui est une évaluation pancanadienne accompagnée de questionnaires contextuels bien structurés, a représenté une mine de renseignements inégalée sur les principaux domaines stratégiques d'intérêt pour les ministères de l'Éducation. Le PPCE permet aux provinces de comparer de façon simple leurs résultats à ceux du reste du Canada. Les données du PPCE fournissent également aux provinces des informations sur le rendement de leur propre système éducatif.

Le CMEC s'engage à promouvoir la recherche ayant trait aux politiques et à l'éducation, et à maintenir le caractère prioritaire de la diffusion des résultats de recherche auprès des responsables de l'élaboration des politiques et des personnes travaillant sur le terrain. Les évaluations du PPCE ont été conçues pour produire des données de rendement à l'échelon pancanadien et à l'échelon provincial/territorial¹⁹. Les données sont également disponibles en fonction de la langue d'enseignement, c'est-à-dire en français ou en anglais, et en fonction du sexe. La taille de l'échantillon est toutefois trop petite pour permettre de produire des résultats

¹⁸ Les scores d'indice associés aux écoles n'ont pas été inclus dans le rapport contextuel.

¹⁹ Aucun territoire n'a participé au PPCE 2016.

fiables à partir d'analyses au sein des sous-catégories d'une province donnée (par exemple, les écoles ou les commissions/conseils/districts scolaires). Pour des raisons de confidentialité, tous les renseignements ayant trait à l'identité des élèves, des écoles et des commissions/conseils scolaires sont supprimés au moment de la préparation des ensembles de données pour l'analyse finale du CMEC.

Aucun ensemble de données permettant d'identifier l'école, la commission, le conseil ou le district scolaire, ou les individus ne peut être rendu disponible.

Les chercheuses et chercheurs qui désirent avoir accès aux ensembles de données du PPCE devront s'engager par écrit à respecter les modalités d'accès ci-dessous.

Conditions et modalités

Le CMEC maintiendra un registre de toutes les demandes d'utilisation des données du PPCE, afin que les provinces/territoires soient tenus régulièrement au courant des recherches entreprises à partir de ces données. Les demandes des chercheuses et chercheurs travaillant dans un domaine autre que celui de l'éducation qui souhaitent utiliser les données du PPCE sont les bienvenues.

En ce qui concerne le registre, les chercheuses et chercheurs souhaitant utiliser les données du PPCE devront fournir les renseignements suivants, lors de leur demande d'accès aux ensembles de données :

- nom(s) et affiliation(s) des chercheuses et chercheurs travaillant sur le projet (par exemple, nom de l'université, du collège, du ministère de l'Éducation, de la commission, du conseil ou du district scolaire, de la fondation de recherche ou encore de l'organisme où la chercheuse ou le chercheur est employé ou pour laquelle ou lequel le travail est effectué);
- coordonnées de la chercheuse ou du chercheur principal (adresse postale, numéro de téléphone, numéro de télécopie, adresse courriel);
- description succincte du projet comprenant les éléments suivants :
 - le ou les buts du projet;
 - la méthodologie proposée pour la recherche;
 - les sources de renseignements et personnes interviewées;
 - les documents du CMEC requis pour effectuer la recherche;
 - les logiciels utilisés (afin de garantir leur compatibilité avec la base de données du PPCE);
 - le plan de diffusion proposé.

En raison des considérations relatives à la taille des échantillons, les chercheuses et chercheurs n'ont pas le droit d'utiliser les données du PPCE en vue de classer les écoles ou les districts/commissions/conseils scolaires, étant donné que de telles comparaisons ne seraient pas valables.

Les demandes d'accès au matériel d'évaluation confidentiel (par exemple, les livrets des épreuves) seront étudiées par le CMEC à la condition stricte que le contenu des livrets et les numéros d'identification ne soient aucunement divulgués dans le rapport à paraître.

La diffusion des résultats est prioritaire dans la recherche sur le PPCE. Le CMEC est particulièrement intéressé par les possibilités de diffusion auprès des responsables de l'élaboration des politiques et auprès des personnes travaillant sur le terrain et accueillerait volontiers des initiatives de recherche comprenant de telles activités. La publication des résultats des recherches relève de la responsabilité des chercheuses et chercheurs, à moins que le CMEC ne décide de jouer un rôle actif dans la diffusion des résultats. Les chercheuses et chercheurs seront responsables de leurs recherches et de leurs conclusions. Les chercheuses et chercheurs devront soumettre au CMEC un rapport sur leurs conclusions ou un exemplaire de leur article avant toute publication ou présentation des conclusions de leurs recherches. Le CMEC distribuera le rapport sur les conclusions aux provinces/territoires membres nommés ou identifiés dans le rapport, en vertu d'une entente de confidentialité, un mois avant la publication ou la parution des conclusions, afin que les provinces/territoires puissent préparer des stratégies de communication en prévision de la publication du rapport par les chercheuses ou chercheurs. Sauf indication contraire, ce rapport sera utilisé par le CMEC uniquement à titre indicatif et le CMEC ne publiera pas le rapport sans le consentement des personnes responsables de la recherche.

La source et le but original de la collecte des données doivent être mentionnés lors de la parution ou de la présentation d'une analyse secondaire des données. Les chercheuses et chercheurs doivent s'engager à ce que les ensembles de données ne soient pas obtenus par d'autres personnes et ce, par quelque moyen que ce soit.

Information pour les chercheuses et chercheurs

Le CMEC s'engage à promouvoir la recherche ayant trait aux politiques et accorde la priorité à la diffusion des résultats de recherche aux responsables de l'élaboration des politiques et aux praticiennes et praticiens. La publication *L'évaluation... ça compte!* est un recueil de notes de recherche axées sur les politiques qui explore les questions d'éducation au Canada à partir des résultats des programmes d'évaluation pancanadiens et internationaux. Ces notes de recherche utilisent des données d'évaluation pertinentes pour répondre à des questions de recherche pressantes sur les enjeux de l'éducation au Canada.

Les chercheuses et chercheurs peuvent obtenir l'accès à des données et à d'autres outils provenant d'évaluations internationales et pancanadiennes de l'apprentissage à partir du site Web du CMEC, à l'adresse suivante :

https://www.cmec.ca/704/Donn%3a9es_d%e2%80%99%3a9valuation_pour_le_milieu_d_e_la_recherche.html.

Bibliographie

- BAKER, F. B., et A. AL-KARNI. « A comparison of two procedures for computing IRT equating coefficients », *Journal of Educational Measurement*, n° 28, 1991, p. 147-162.
- BENJAMIN, R. « Reconstructing readability: Recent developments and recommendations in the analysis of text difficulty », *Educational Psychology Review*, vol. 24, n° 1, 2012, p. 63-88.
- BRIGGS, D. « An introduction to Multidimensional IRT », communication présentée à UC Berkeley, avril 2008. Sur Internet :
http://www.powershow.com/view/3c4039-MmRjY/An_Introduction_to_Multidimensional_IRT_Derek_Briggs_April_powerpoint_ppt_presentation
- CAI, L. « Metropolis-Hastings Robbins-Monro algorithm for confirmatory item factor analysis », *Journal of Educational and Behavioral Statistics*, vol. 35, n° 3, 2010, p. 307-335.
- CHEN, W.-H., et D. THISSEN. « Local dependence indexes for item pairs using item response theory », *Journal of Educational and Behavioral Statistics*, vol. 22, n° 3, 1997, p. 265-289.
- CIZEK, G. J., et M. B. BUNCH. *Standard setting: A guide to establishing and evaluating performance standards on tests*, Thousand Oaks, CA, SAGE Publications Ltd., 2007.
- CONSEIL DES MINISTRES DE L'ÉDUCATION (CANADA). *Cadre commun de résultats d'apprentissage des sciences de la nature M à 12 – Protocole pancanadien pour la collaboration en matière de programmes scolaires*, Toronto, CMEC, 1997. Sur Internet :
<http://science.cmec.ca/framework/>
- CONSEIL DES MINISTRES DE L'ÉDUCATION (CANADA). *Étude documentaire des programmes d'études des sciences et des cadres d'évaluation*, Toronto, CMEC, 2005 [inédit].
- CONSEIL DES MINISTRES DE L'ÉDUCATION (CANADA). *PPCE de 2010 – Programme pancanadien d'évaluation*, Toronto, CMEC, 2011.
- CONSEIL DES MINISTRES DE L'ÉDUCATION (CANADA). *Cadre d'évaluation du PPCE 2016*, Toronto, CMEC, 2016.
- DE AYALA, R. J. *The theory and practice of item response theory*, New York, Guilford Press, 2009.
- DE AYALA, R. J., B. S. PLAKE et J. C. IMPARA. « The impact of omitted response on the accuracy of ability estimation in item response theory », *Journal of Educational Measurement*, vol. 38, n° 3, 2001, p. 213-234.
- DE CHAMPLAIN, A. F., et M. E. GESSAROLI. « Assessing the dimensionality of item response matrices with small sample size and short test lengths », *Applied Measurement in Education*, n° 11, 1998, p. 231-253.
- FENSHAM, P., et W. HARLEN. « School science and public understanding of science », *International Journal of Science Education*, vol. 21, n° 7, 1999, p. 755-763.

- FINCH, H. « Estimation of item response theory parameters in the presence of missing data », *Journal of Educational Measurement*, vol. 45, n° 3, p. 225-245.
- FINCH, H., et B. HABING. « Comparison of NOHARM and DETECT in item cluster recovery: Counting dimension and allocating items », *Journal of Educational Measurement*, n° 42, p. 149-169.
- FRASER, C., et R. P. McDONALD. *NOHARM: A window program for fitting both unidimensional and multidimensional normal ogive models of latent trait theory* [logiciel], Welland (Ontario), Niagara College, 2003. Sur Internet : <http://noharm.software.informer.com>
- FOY, P., B. BROSSMAN et J. GALIA. « Scaling the TIMSS and PIRLS 2011 achievement data », dans M. O. Martin et I. V. S. Mullis (éd.), *Methods and procedures in TIMSS and PIRLS 2011*, Chestnut Hill, MA, TIMSS & PIRLS International Study Center, Boston College, 2012. Sur Internet : http://timssandpirls.bc.edu/methods/pdf/TP11_Scaling_Achievement.pdf
- HIDI, S., et D. BERNDORFF. « Situational interest and learning », dans L. Hoffmann, A. Krapp, K.A. Renniger et J. Baumert (éd.), *Interest and Learning*, Kiel (Allemagne), Institut des sciences de l'éducation de l'Université de Kiel, 1998.
- HOY, A. W. « Changes in teacher efficacy during the early years of teaching », communication présentée à l'assemblée générale annuelle de l'*American Educational Research Association* (association américaine de la recherche sur l'éducation), Nouvelle-Orléans, avril 2000.
- JOHNSON, M. S. « Marginal maximum likelihood estimation of item response models », *Journal of Statistical Software*, vol. 20, n° 10, 2007, p. 1-19.
- KANDEL, L., et A. MOLES. « Application de l'indice de Flesch à la langue française », *Cahiers d'études de radio-télévision*, n° 19, 1958, p. 253-274.
- KIM, S., et M. J. KOLEN. « Robustness to format effects of IRT linking methods for mixed-format tests », *Applied Psychological Measurement*, vol. 19, n° 4, 2006, p. 357-381.
- KLARE, G. R. « The formative years », dans B. L. Zakaluk et S. J. Samuels (éd.), *Readability, its past, present and future*, Newark, Delaware, International Reading Association, 1988, p. 14-34.
- KNOL, W. R., et M. P. F. BERGER. « Empirical comparison between factor analysis and multidimensional item response models », *Multivariate Behavioral Research*, n° 26, 1991, p. 457-477.
- LAVEAULT, D., et J. GRÉGOIRE. *Introduction aux théories des tests en psychologie et en éducation*, 2^e éd., Bruxelles, De Boeck, 2002.
- LORD, F. M. « Estimation of latent ability and item parameters when there are omitted responses », *Psychometrika*, n° 39, 1974, p. 247-264.
- LORD, F. M. « Maximum likelihood estimation of item parameters when some responses are omitted », *Psychometrika*, n° 48, 1983, p. 477-482.
- LUDLOW, L. H., et M. O'LEARY. « Scoring omitted and not-reached items: Practical data analysis implications », *Educational and Psychological Measurement*, n° 59, 1999, p. 615-630.

- McDONALD, R. P. *Nonlinear factor analysis*, coll. « Psychometric Monographs », n° 15, Richmond, VA, Psychometric Corporation, 1967. Sur Internet :
<https://www.psychometricsociety.org/sites/default/files/pdf/MN15.pdf>
- MARTIN, M. O., I. V. S. MULLIS, P. FOY, B. BROSSMAN et G. M. STANCO. « Estimating linking error in PIRLS », *IERI Monograph Series: Issues and Methodologies in Large-Scale Assessments*, n° 5, 2012, p. 35-47. Sur Internet :
http://www.ierinstitute.org/fileadmin/Documents/IERI_Monograph/IERI_Monograph_Volume_05_Chapter_2.pdf
- MURAKI, E., et G. ENGELHARD. « Examining differential item functioning with BIMAIN », communication présentée à l'assemblée générale annuelle de l'American Educational Research Association, San Francisco, avril 1989.
- NATIONAL COUNCIL OF TEACHERS OF MATHEMATICS. *Principles and Standards for School Mathematics*, Reston, VA, NCTM, 2000.
- O'GRADY, K., K. FUNG, L. SERVAGE et G. KHAN. *PPCE 2016 : Rapport de l'évaluation pancanadienne en lecture, en mathématiques et en sciences*, Toronto, Conseil des ministres de l'Éducation (Canada), 2018.
- ORGANISATION DE COOPÉRATION ET DE DÉVELOPPEMENT ÉCONOMIQUES (OCDE). *PISA 2009 Technical Report*, Paris, PISA, Éditions OCDE, 2012. Sur Internet :
<http://dx.doi.org/10.1787/9789264167872-en>
- ORGANISATION DE COOPÉRATION ET DE DÉVELOPPEMENT ÉCONOMIQUES (OCDE). *PISA 2006 – Les compétences en sciences, un atout pour réussir*, Paris, PISA, Éditions OCDE, 2006.
- OSBORNE, J., S. SIMON et S. COLLINS. « Attitudes towards science: A review of the literature and its implications », *International Journal of Science Education*, vol. 25, n° 9, 2003, p. 1049-1079.
- RIGGS, I., et L. ENOCHS. « Towards the development of an elementary teacher's science teaching efficacy belief instrument », *Science Education*, n° 74, 1990, p. 625-637.
- WANG, M. C., G. D. HAERTEL et H. J. WALBERG. « Synthesis of research: What Helps Students Learn? », *Educational Leadership*, décembre 1993/janvier 1994, p. 74-79.
- WANG, M. C., G. D. HAERTEL et H. J. WALBERG. « Toward a knowledge base for school learning », *Review of Educational Research*, vol. 63, n° 3, 1993, p. 249-294.
- WANG, M. C., G. D. HAERTEL et H. J. WALBERG. « What influences learning? A content analysis of review literature », *Journal of Educational Research*, vol. 84, n° 1, 1990, p. 30-43.
- YEN, W. M. « Scaling performance assessments: Strategies for managing local item dependence », *Journal of Educational Measurement*, n° 30, 1993 p. 187-213.
- ZHANG, B., et C. M. WALTER. « Impact of missing data on person-model fit and person trait estimation », *Applied Psychological Measurement*, vol. 32, n° 6, 2008, p. 466-479.

Programme pancanadien d'évaluation (PPCE)

Plan d'échantillonnage et pondération

Préparé pour le CMEC par :
Groupe de consultation statistique
Division de la coopération internationale et des méthodes statistiques institutionnelles
(DCIMSI)
Statistique Canada
Novembre 2015 (plan d'échantillonnage)
Novembre 2016 (pondération)

Table des matières

1. Introduction	3
2. Plan d'échantillonnage.....	3
2.1 Populations cibles et populations observées.....	4
2.2 Base de sondage	5
2.2.1 Premier degré de la base de sondage (liste des écoles à l'intérieur du champ)	5
2.2.2 Deuxième degré de la base de sondage (liste des classes à l'intérieur du champ)	7
2.3 Stratification.....	9
2.4 Répartition de la taille de l'échantillon.....	9
2.5 Sélection de l'échantillon.....	11
3. Pondérations d'échantillonnage	12
4. Rajustements de la pondération de l'échantillonnage pour la non-réponse	13
5. Pondérations bootstrap	17
Tableau A – Nouveaux codes de participation	20
Tableau B – Résumé des rajustements manuels des numéros d'identification des écoles	21
Références	24

1. Introduction

Le Conseil des ministres de l'Éducation (Canada) [CMEC] est responsable de la mise en œuvre de la quatrième édition du Programme pancanadien d'évaluation (PPCE). Cette édition du PPCE, qui est conçu pour évaluer les compétences des élèves de 8^e année (2^e secondaire au Québec) dans les domaines des sciences, des mathématiques et de la lecture, a été déployée au printemps 2016.

Le CMEC a retenu les services de Statistique Canada pour concevoir et mettre en œuvre la stratégie d'échantillonnage aléatoire pour le PPCE 2016. Le présent document vise à présenter en détail tous les aspects du plan d'échantillonnage. Plus précisément, ce document couvre des sujets tels que : la population cible et la population observée; la base de sondage; la stratification, la taille de l'échantillon et les méthodes de sélection de l'échantillon; ainsi que la méthodologie proposée et le calendrier pour les activités d'échantillonnage.

2. Plan d'échantillonnage

Le *plan d'échantillonnage* fait référence à la constitution de l'échantillon et à la façon dont celui-ci est obtenu. Le plan d'échantillonnage est un ensemble de spécifications qui décrivent la population cible et la population observée, la base de sondage, la stratification, la taille de l'échantillon et les méthodes de sélection de l'échantillon. Ces caractéristiques de conception sont décrites en détail dans les sous-sections 2.1 à 2.5.

Comme il est indiqué ci-dessus, l'objectif du PPCE est d'évaluer les compétences des élèves de 8^e année (2^e secondaire au Québec) dans les 10 provinces du Canada. Puisqu'il n'existe pas de liste unique de tous les élèves de 8^e année (2^e secondaire) au Canada, l'échantillon du PPCE a été sélectionné au moyen d'un plan d'échantillonnage probabiliste stratifié à deux degrés. Ainsi, l'échantillon d'élèves de 8^e année (2^e secondaire) a été sélectionné en deux étapes successives.

À la première étape, un échantillon aléatoire d'écoles (les unités primaires d'échantillonnage, ou UPE) a été sélectionné à partir d'une liste d'écoles à l'intérieur du champ fournie par le CMEC. Toutes les classes de 8^e année (2^e secondaire) à l'intérieur du champ dans les écoles sélectionnées ont alors été dénombrées. La deuxième étape de la sélection de l'échantillon a consisté à sélectionner aléatoirement une classe de 8^e année (2^e secondaire) par école sélectionnée. Les classes de 8^e année (2^e secondaire) ont par conséquent été les unités secondaires d'échantillonnage (USE). Tous les élèves à l'intérieur du champ des classes de 8^e année (2^e secondaire) sélectionnées ont participé à l'évaluation du PPCE.

2.1 Populations cibles et populations observées

La définition de la population à partir de laquelle un échantillon est sélectionné est une étape essentielle de l'élaboration d'un plan d'échantillonnage valable. Une bonne définition facilite le processus d'échantillonnage et prévient les ambiguïtés. Les populations peuvent être appelées « populations cibles » ou « populations observées ».

La *population cible* est l'ensemble des éléments pour lesquels des renseignements sont souhaités et des estimations doivent être produites. La *population observée* est un sous-ensemble de la population cible qui est obtenu à partir de la base de sondage et d'autres considérations pratiques. La population observée peut ne pas être exactement la même que la population cible, mais, idéalement, elle devrait être très similaire.

La population cible pour cette enquête est composée de tous les élèves de 8^e année (2^e secondaire) dans les 10 provinces du Canada. Les écoles qui ne sont pas financées par les provinces sont exclues de la population cible. Les catégories suivantes d'élèves sont également exclues de la population cible :

- élèves ayant une déficience fonctionnelle ou intellectuelle;
- élèves qui fréquentent des écoles fédérales ou internationales;
- élèves qui sont au Canada depuis moins de deux ans ou qui ne parlent ni le français ni l'anglais.

Pour des raisons pratiques, la population observée pour la présente étude exclut les écoles isolées géographiquement et les écoles ayant moins de cinq élèves de 8^e année (2^e secondaire). Notre analyse indique que ces exclusions correspondent à au plus 0,7 p. 100 et 2,1 p. 100 respectivement de la population cible à l'échelle pancanadienne et à l'échelle provinciale. Bien qu'elles n'aient pas explicitement été exclues de la présente étude, les écoles sélectionnées qui avaient précédemment participé à d'autres évaluations ont pu être remplacées par des écoles de remplacement sélectionnées à la discrétion des administratrices et administrateurs du PPCE.

Deux types d'unités doivent généralement être distingués dans la population observée : les unités d'échantillonnage et les unités déclarantes. Pour la présente étude, il existe deux types d'unités d'échantillonnage. À la première étape, les unités d'échantillonnage sont les écoles. À la deuxième étape, les unités d'échantillonnage sont les classes composées d'élèves de 8^e année (2^e secondaire). Pour la présente étude, les unités déclarantes sont les élèves de 8^e année (2^e secondaire) dans les classes sélectionnées.

La période de référence est définie comme la période à laquelle les données se rapportent. Pour la présente étude, la période de référence est définie comme la période pendant laquelle les évaluations du PPCE se sont déroulées, soit au printemps 2016.

2.2 Base de sondage

Le terme *base de sondage* englobe toute liste, tout document ou tout appareil qui délimite, définit les éléments de la population observée et y donne accès. La base contient toutes les unités qui composent la population à partir de laquelle l'échantillon est prélevé.

La qualité de la base détermine la couverture de l'enquête et a une incidence sur l'efficacité du plan d'échantillonnage. L'omission et l'inclusion erronées d'unités dans la base entraînent un sous-dénombrement et un surdénombrement, respectivement, dans l'étude.

Deux bases de sondage sont utilisées pour cette étude – une pour chaque étape de la sélection de l'échantillon. Les sous-sections 2.2.1 et 2.2.2 fournissent de plus amples détails sur les bases de sondages pour la première et la deuxième étape de la sélection de l'échantillon.

2.2.1 Premier degré de la base de sondage (liste des écoles à l'intérieur du champ)

Pour la première étape de la sélection de l'échantillon, le CMEC a obtenu des provinces la liste de toutes les écoles à l'intérieur du champ. Cette liste a été fournie à Statistique Canada pour les besoins du premier degré de la base de sondage. Elle comporte un numéro d'identification des écoles, la province, la langue du conseil ou de la commission scolaire (français ou anglais), le nombre d'élèves de 8^e année (2^e secondaire) dans l'école, le nombre de classes de 8^e année (2^e secondaire) et un indicateur d'exemption précisant si l'école devrait être exclue de la couverture de l'enquête ainsi que les raisons de l'exemption (Tableau 2.1).

Tableau 2.1 : Variables importantes pour la première étape de la sélection de l'échantillon

Variable	Description
ID_ÉCO_PRO	Numéro d'identification de l'école
PROVINCE	Province
CONSEILSCOLAIRE	Langue du conseil ou de la commission scolaire (français ou anglais)
ÉCOLE_8ANNÉE	Nombre d'élèves de 8 ^e année (2 ^e secondaire)
CLASSES_ÉCOLE	Nombre de classes de 8 ^e année (2 ^e secondaire)
EXEMPTION	Indicateur d'exemption de l'école

La section 2.1 a défini les quelques types d'écoles qui ne sont pas couvertes par cette étude. Pour enlever ces écoles de la base d'échantillonnage avant l'échantillonnage, l'indicateur d'exemption a été utilisé. Le tableau suivant (Tableau 2.2) fournit de l'information pour les types d'écoles exclues de cette étude.

Tableau 2.2 : Documentation sur les indicateurs d'exemption

Exemption	Description	Commentaire
1	École spéciale	Exclue
2	École au sein d'une autre province	Exclue
3, 9	École isolée géographiquement	Exclue
4	École déjà soumise à une autre évaluation	Non exclue. Peut être remplacée par une école de remplacement.
5, 7, 8	École fédérale/internationale	Exclue
6	École comptant moins de 9 élèves	Non exclue. ÉCOLE_8ANNÉE <5 utilisée en remplacement.
10	École non financée	Exclue
11	École fermée	Exclue

Pour vérifier la couverture de la base du PPCE 2016, nous avons comparé le nombre d'élèves de 8^e année (2^e secondaire) déclaré dans la base avec les projections démographiques du recensement pour les élèves de 13 ans. Nous avons également comparé ces chiffres à ceux observés dans la base du PPCE 2013 (Tableau 2.3). À la lumière de ces résultats, nous avons noté que le nombre d'élèves de 8^e année (2^e secondaire) dans la base de 2016 correspond bien aux chiffres de 2013 ainsi qu'aux projections du recensement. En 2015, à l'échelle pancanadienne, nous observons que la base du PPCE offre une bonne couverture de la population, les chiffres du cadre correspondant presque exactement aux projections de la population, les chiffres de la base correspondant presque exactement aux projections démographiques. À l'échelle provinciale, le pourcentage de la population couverte se situe en moyenne autour de 95 p. 100. Cependant, ce chiffre ne s'élève qu'à 89,4 p. 100 au Manitoba. Étant donné ces comparaisons, nous sommes d'avis que le cadre du PPCE est fiable et concluons qu'il est de bonne qualité.

Tableau 2.3 : Couverture du cadre du PPCE

PROVINCE	Projections de la population* (élèves de 13 ans)		Base		Proportion de la population	
	2013	2015	2013	2015	2013	2015
Colombie-Britannique	47 129	45 848	46 895	45 576	99,5 %	99,4 %
Alberta	45 386	45 354	40 094	46 578	88,3 %	102,7 %
Saskatchewan	13 677	13 371	12 598	12 335	92,1 %	92,3 %
Manitoba	16 289	15 852	14 451	14 174	88,7 %	89,4 %
Ontario	153 091	147 322	145 756	146 765	95,2 %	99,6 %
Québec	79 921	78 354	85 435	81 563	106,9 %	104,1 %
Nouveau-Brunswick	8 010	7 543	7 976	7 327	99,6 %	97,1 %
Nouvelle-Écosse	9 785	9 146	9 792	8 674	100,1 %	94,8 %
Île-du-Prince-Édouard	1 657	1 527	1 487	1 433	89,7 %	93,8 %
Terre-Neuve-et-Labrador	5 436	5 221	5 441	5 193	100,1 %	99,5 %
CANADA	380 381	369 538	369 925	369 618	97,3 %	100,0%

*Source : Statistique Canada, Tableau 051-0001 – Données dérivées des estimations de la population, selon le groupe d'âge et le sexe au 1^{er} juillet, Canada, provinces et territoires – annuel (personnes sauf indication contraire)

2.2.2 Deuxième degré de la base de sondage (liste des classes à l'intérieur du champ)

Une fois qu'un échantillon d'écoles est sélectionné, le CMEC travaille avec les provinces pour dénombrier toutes les classes de 8^e année (2^e secondaire) au sein des écoles sélectionnées. Comme les élèves de 8^e année (2^e secondaire) peuvent changer de classes au cours de la journée, il est recommandé que l'énumération vise les classes-foyers de 8^e année (2^e secondaire) ou des classes dans une matière que tous les élèves de 8^e année (2^e secondaire) sont obligés de suivre (p. ex., mathématiques ou langue première). Le Tableau 2.4 présente des variables importantes requises pour la deuxième étape de la sélection de l'échantillon.

Tableau 2.4 : Variables importantes pour la deuxième étape de la sélection de l'échantillon

Variable	Description
ID DE L'ÉCHANTILLON	Numéro d'identification de l'école du dossier d'échantillon de l'école
ID DE LA CLASSE	Au sein de chaque école, un seul numéro

	d'identification de la classe
NOM DE L'ENSEIGNANTE OU ENSEIGNANT	Nom de l'enseignante ou enseignant, redondant si l'ID DE LA CLASSE est unique
ÉLÈVES DE 8 ^E ANNÉE/ 2 ^E SECONDAIRE	Nombre d'élèves de 8 ^e année (2 ^e secondaire) dans chaque classe
EXEMPTION	Indicateur d'exemption de la classe

Il se peut que certaines écoles n'aient pas de classes bien définies. Dans ce cas, nous recommandons que tous les élèves de 8^e année (2^e secondaire) soient dénombrés. Il se peut aussi que certaines écoles n'aient pas de niveaux scolaires clairement définis. Dans ce cas, nous recommandons que tous les élèves dont la date de naissance se situe entre le 1^{er} janvier et le 31 décembre 2002 soient dénombrés.

La section 2.1 a présenté les élèves qui ne sont pas couverts par cette étude. Ces élèves, ou classes d'élèves, ont été enlevés de la base avant l'échantillonnage. Une classe entière peut être exclue si tous les élèves se trouvent dans une catégorie pour laquelle les élèves sont exclus; ces trois catégories, ainsi que les codes de catégories connexes, sont indiqués ci-dessous.

- F Exemption pour cause d'incapacités fonctionnelles – Un élève qui a une incapacité physique et est incapable de travailler dans le contexte du test du PPCE, même avec une ou plusieurs des sept adaptations autorisées, doit être exempté. Les élèves ayant une incapacité fonctionnelle, mais qui sont capables de participer, devraient participer à l'évaluation.
- I Exemption pour cause d'incapacités intellectuelles ou de problèmes socioaffectifs – Un élève qui est considéré, selon l'opinion professionnelle de la direction de l'école ou d'autres membres qualifiés du personnel, comme ayant une incapacité intellectuelle ou des problèmes socioaffectifs ou dont les tests ont montré qu'il avait une telle incapacité ou de tels problèmes doit être exempté. Cette catégorie inclut les élèves qui sont incapables, sur le plan mental ou affectif, de suivre même les instructions générales du test.
- N Exemption en raison de la langue (locutrices ou locuteurs non natifs) – Cette exemption ne s'applique qu'aux élèves dont la langue première n'est ni le français ni l'anglais. Dans les évaluations de grande envergure, les écoles peuvent examiner le cas des élèves qui sont au Canada depuis moins de deux ans.

Toutes les exemptions à l'échelle de la classe doivent être approuvées par le CMEC.

2.3 Stratification

La *stratification* est un moyen d'organiser la base de sondage pour la rendre plus précise pour une taille d'échantillon fixe. La stratification peut aussi être utilisée pour garantir qu'un nombre minimal d'unités soit obtenu ou que les exigences relatives à la précision pour certains groupes de population soient respectées. Les strates sont exhaustives et sont des groupes d'écoles mutuellement exclusifs. Chaque école ne figure que dans une strate. La taille de l'échantillon total est répartie parmi les strates. Chaque strate est échantillonnée de façon indépendante.

Il est important pour le CMEC de publier des statistiques fiables à l'échelle pancanadienne et provinciale et à l'égard de la langue des conseils et commissions scolaires au sein des provinces. Pour garantir un échantillon suffisamment important dans ces volets, les strates du PPCE sont définies comme la classification croisée de la province selon la langue du conseil ou de la commission scolaire.

2.4 Répartition de la taille de l'échantillon

La *répartition de la taille de l'échantillon* est un processus de compromis dans lequel les exigences en matière de précision des estimations sont pondérées avec diverses contraintes opérationnelles comme le temps, le budget et les ressources disponibles.

Pour le PPCE 2016, nous avons commencé le processus de répartition de la taille de l'échantillon en considérant la stratégie utilisée en 2013. En étudiant cette stratégie de répartition, nous avons déterminé que son rendement sera très satisfaisant pour ce qui est de satisfaire aux exigences requises en matière de précision. Cette répartition a aussi l'avantage d'être approuvée au préalable par les provinces. Pour ces raisons, la répartition de la taille de l'échantillon du PPCE 2016 est essentiellement la même que celle de l'évaluation précédente. Le tableau ci-dessous (Tableau 2.5) montre la répartition de la taille de l'échantillon de 2016. En observant ces résultats, nous constatons qu'un recensement des écoles sera fait dans 13 de 20 strates.

Tableau 2.5 : Répartition de la taille de l'échantillon du PPCE 2016

Strate	Population à l'intérieur du champ		Échantillon			
	Écoles	Élèves	Écoles	Élèves*	Toutes les écoles	Tous les élèves
Alberta – anglophone	725	45 622	150	2 990		
Alberta – francophone	20	426	20	299	Oui	
Colombie-Britannique – anglophone	462	44 887	150	3 011		
Colombie-Britannique – francophone	12	262	12	171	Oui	
Manitoba – anglophone	321	13 539	150	2 846		
Manitoba – francophone	19	369	19	267	Oui	
Nouveau-Brunswick – anglophone	83	5 122	83	1 516	Oui	
Nouveau-Brunswick – francophone	61	2 195	61	1 164	Oui	
Terre-Neuve-et-Labrador – anglophone	114	5 039	114	1 865	Oui	
Terre-Neuve-et-Labrador – francophone	3	23	3	23	Oui	Oui
Nouvelle-Écosse – anglophone	116	8 333	116	2 480	Oui	
Nouvelle-Écosse – francophone	11	329	11	184	Oui	
Ontario – anglophone	2 930	139 230	150	2 713		
Ontario – francophone	151	6 367	125	2 094		
Île-du-Prince-Édouard – anglophone	22	1 369	22	438	Oui	
Île-du-Prince-Édouard – francophone	3	58	3	58	Oui	Oui
Québec – anglophone	89	7 425	89	1 801	Oui	
Québec – francophone	480	71 909	150	3 361		
Saskatchewan – anglophone	440	11 992	186	3 143		
Saskatchewan – francophone	5	79	5	79	Oui	Oui
Total Canada	6 067	364 575	1 619	30 503	13/20	3/20

* Taille de l'échantillon prévue

2.5 Sélection de l'échantillon

La *sélection de l'échantillon* fait référence au processus utilisé pour obtenir l'échantillon de l'enquête à partir de la population observée. L'objectif de la présente sous-section est de documenter le processus de sélection de l'échantillon pour la première et la deuxième étape de la sélection de l'échantillon.

Pour la première étape de la sélection de l'échantillon (la sélection des écoles), deux méthodes ont été utilisées. Pour 13 des 20 strates, un recensement des écoles a été fait. Pour ces strates, la méthode de sélection de l'échantillon était simple, puisque toutes les écoles ont été sélectionnées pour l'échantillon. Pour les sept strates restantes, la méthode de sélection de l'échantillon utilisée a été l'échantillonnage systématique (ES), dans le cadre duquel les écoles ont d'abord été classées en ordre décroissant selon le nombre d'élèves de 8^e année (2^e secondaire) au sein de chaque école. Ce classement est une façon simple de veiller à ce que l'échantillon des écoles soit réparti proportionnellement à la taille des écoles. Cette méthode comporte deux avantages. Premièrement, elle est avantageuse sur le plan administratif, puisque des écoles de remplacement sont situées directement au-dessus et au-dessous des écoles sélectionnées dans la liste classée des écoles. Le deuxième avantage de cette méthode de sélection de l'échantillon est qu'elle utilise l'information auxiliaire disponible à l'étape de sélection de l'échantillon. Par conséquent, si les estimations d'intérêt sont corrélées avec la taille des écoles, les estimations qui en résultent peuvent avoir une variabilité d'échantillonnage moindre qu'un plan d'échantillonnage qui n'intègre pas cette information auxiliaire. Le désavantage de cette méthode est que, si la taille des écoles n'est pas corrélée avec les estimations d'intérêt, les estimations qui en résultent peuvent avoir une plus grande variabilité d'échantillonnage qu'un plan d'échantillonnage qui ne tient pas compte de ces données auxiliaires.

Le CMEC doit déployer tous les efforts pour confirmer la participation du plus grand nombre possible d'écoles échantillonnées. Cette participation est importante pour minimiser la possibilité de biais de non-réponse. Après avoir communiqué avec toutes les écoles échantillonnées, le CMEC devra ensuite communiquer avec les écoles de remplacement pour les écoles échantillonnées qui ne participeront pas. Chaque école échantillonnée ne participant pas doit être remplacée, si possible, par la première école de remplacement. La deuxième école de remplacement ne devrait être utilisée que si l'école de l'échantillon correspondante et la première école de remplacement ne participent pas toutes les deux. ***Si l'école de l'échantillon originale n'est pas admissible, ou qu'il s'agit d'un type d'école appartenant aux écoles exemptées définies, ou si l'école a été fermée, alors aucune école de remplacement ne devrait être utilisée.*** De plus, une école avec peu, mais suffisamment d'élèves de 8^e année (2^e secondaire) ne devrait pas être remplacée simplement parce que le nombre d'élèves pourrait être inférieur à celui escompté.

En général, les écoles de remplacement sont utilisées pour contrer la non-réponse (c.-à-d., si et seulement si la première école refuse de participer). Dans la méthodologie de l'enquête, la non-réponse est traitée de deux façons. La première est au moyen du rajustement de la

pondération, lorsque les écoles répondantes peuvent représenter les écoles non répondantes. La deuxième façon est appelée « imputation ». Dans ce cas, les valeurs des réponses sont attribuées aux écoles non répondantes au moyen des données recueillies à partir des écoles répondantes. Aux termes de ce cadre, les écoles de remplacement sont considérées comme une forme d'imputation. Dans le cas où des écoles en dehors du champ sont choisies, nous ne souhaitons pas attribuer cette pondération à d'autres écoles ou les considérer avec l'imputation, puisque ces écoles en dehors du champ sélectionnées représentent effectivement toutes les autres écoles en dehors du champ dans la base de sondage.

Bref, les écoles de remplacement sont utilisées pour traiter la non-réponse. Les écoles en dehors du champ ne reçoivent aucun traitement, puisqu'elles reflètent le fait que la base n'est pas à jour. Cette perte de taille de l'échantillon est le prix à payer pour avoir un cadre qui n'est pas à jour.

Pour la deuxième étape de la sélection de l'échantillon (c.-à-d., la sélection des classes), trois méthodes de sélection ont été utilisées. Pour les écoles qui comptaient plus de 20 élèves et qui ont été en mesure de dénombrer toutes les classes de 8^e année (2^e secondaire), une classe de 8^e année (2^e secondaire) par école a été sélectionnée au moyen d'un échantillon aléatoire simple (EAS). Tous les élèves à l'intérieur du champ dans les classes sélectionnées ont fait l'objet de l'évaluation. Pour les écoles qui avaient plus de 20 élèves, mais qui n'ont pas été en mesure de dénombrer les classes de 8^e année (2^e secondaire), un dénombrement de tous les élèves de 8^e année (2^e secondaire) a été effectué, et un EAS de 20 élèves a été sélectionné. Pour les écoles comptant 20 élèves de 8^e année (2^e secondaire) ou moins, un recensement des élèves dans ces écoles a été effectué.

3. Pondérations d'échantillonnage

La sous-section 2.5 décrit en détail les méthodes de la première et de la deuxième étape de la sélection de l'échantillon. La présente section vise à spécifier précisément les pondérations du plan d'échantillonnage qui découlent des méthodes de sélection de l'échantillon susmentionnées.

La première étape de la sélection de l'échantillon (la sélection des écoles) consiste à sélectionner chaque école avec une certaine probabilité d'inclusion. L'inverse de cette probabilité d'inclusion est la pondération du plan d'échantillonnage de l'école. Pour définir les pondérations du plan, certaines notations sont requises. Si l'on considère que l'ensemble des écoles de la population observée est de taille N , et que h représente les strates desquelles il y a $H = 20$. Considérons que n_h est la taille de l'échantillon attribuée à la strate h . Considérons que X_i est le nombre total d'élèves de 8^e année (2^e secondaire) dans l'école. Alors, la pondération du plan d'échantillonnage ($d_i^{école}$) pour l'école sélectionnée i est donnée comme :

$$d_i^{écoles} = \begin{cases} \frac{\sum_{i \in h} X_i}{n_h X_i}, & \text{si } n_h < N_h \\ 1, & \text{si } n_h = N_h. \end{cases}$$

Pour tenir compte de la deuxième étape de la sélection de l'échantillon (la sélection des classes dans les écoles), nous appliquons le rajustement de la pondération suivant (raj_i^{classe}) à la pondération de l'école ci-dessus :

$$raj_i^{classe} = \begin{cases} m_i, & \text{si } X_i > 20 \text{ et toutes les classes sont dénombrées} \\ \frac{X_i}{20}, & \text{si } X_i > 20 \text{ et les élèves sont dénombrés} \\ 1, & \text{si } X_i \leq 20 \end{cases}$$

où m_i est le nombre total de classes de 8^e année (2^e secondaire) dans l'école i . La pondération du plan d'échantillonnage des élèves ($d_i^{élèves}$) est définie comme :

$$d_i^{élèves} = d_i^{écoles} \times raj_i^{classe}.$$

Pour arriver aux pondérations des estimations – ces pondérations devant être utilisées dans la production des estimations finales – d'autres rajustements de la pondération pour la non-réponse des écoles et la non-réponse des élèves ont été appliqués. Pour faciliter le calcul des rajustements de la pondération de la non-réponse à l'échelle des élèves, il est nécessaire que le CMEC fournisse à Statistique Canada le nombre total d'élèves à l'intérieur du champ par classe sélectionnée, ou par école si les classes ne peuvent pas être dénombrées.

Le CMEC, en collaboration avec les écoles et les provinces, effectue les évaluations du PPCE. Une fois de plus, tous les élèves sélectionnés ou tous les élèves à l'intérieur du champ dans les classes sélectionnées ont été soumis à l'évaluation. Puisqu'il y avait quatre livrets distincts, nous avons recommandé que la distribution des livrets soit effectuée à partir d'un seul point de départ aléatoire dans chaque province. De plus, nous avons recommandé que la distribution des livrets soit coordonnée entre les écoles. Ainsi, l'école numéro 1 a commencé à distribuer les livrets au point de départ aléatoire, et l'école numéro 2 a commencé à distribuer les livrets où l'école numéro 1 s'est arrêtée, et ainsi de suite. Cette méthode a permis d'assurer la distribution équitable de chaque livret.

4. Rajustements de la pondération de l'échantillonnage pour la non-réponse

Les pondérations des écoles pour les écoles participantes ont été gonflées pour tenir compte des écoles qui n'ont pas répondu. Les écoles qui ont été exemptées conservent leur pondération, et représentent toutes les autres écoles exemptées qui existent dans la population.

En pratique, le rajustement pour la non-réponse des écoles a été effectué en deux étapes. Dans une première étape, des classes ont été sélectionnées et des écoles ont pu décider de ne pas

participer, ce qui correspond à la non-réponse initiale. Dans d'autres cas, après cette première étape, il s'est avéré que d'autres écoles n'avaient pas participé à l'évaluation après que les données ont été retournées.

Voici de nouveau la pondération de l'école initiale ($d_i^{école}$, {**pondération_plan d'échantillonnage_école**}¹) :

$$d_i^{école} = \begin{cases} \frac{\sum_{i \in h} X_i}{n_h X_i}, & \text{si } n_h < N_h \\ 1, & \text{si } n_h = N_h. \end{cases}$$

Il y a ensuite le facteur de rajustement initial pour la non-réponse de l'école ($nr_initiale_i^{école}$ {rajust_nr_école}) :

$$nr_initiale_i^{école} = \frac{\sum_{i \text{ écoles participantes}} d_i^{école} + \sum_{i \text{ écoles non répondantes}} d_i^{école}}{\sum_{i \text{ écoles participantes}} d_i^{école}}.$$

La pondération intermédiaire de l'école ($p_i^{école}$ {**pondération_école**}) est obtenue au moyen du calcul suivant :

$$p_i^{école} = d_i^{école} * nr_initiale_i^{école}.$$

Un rajustement subséquent pour la non-réponse a été créé pour la non-réponse de l'école qui a été obtenue, souvent en raison de grèves du zèle ou d'autres interruptions imprévues, avec ce facteur ($nr_sub_i^{école}$ {rajust_nr_école}) :

$$nr_rem_i^{école} = \frac{\sum_{i \text{ écoles participantes}} p_i^{école} + \sum_{i \text{ écoles non répondantes}} p_i^{écoles}}{\sum_{i \text{ écoles participantes}} p_i^{école}}.$$

La pondération finale de l'école ($p_{finale_i}^{école}$ {**pondération_finale_école**}) étant obtenue comme suit :

$$p_{finale_i}^{école} = nr_rem_i^{école} * p_i^{école} = nr_rem_i^{école} * nr_initiale_i^{école} * d_i^{école}.$$

Il importe de noter que les classes exemptées ne sont pas incluses dans le rajustement. Les classes exemptées reçoivent une pondération qui est conservée pour représenter toutes les autres classes exemptées qui ne sont pas trouvées, en raison d'une base de sondage qui n'est pas à jour. Ces pondérations sont conservées pour que les taux d'exemption puissent être

¹ Les noms des variables correspondantes dans les dossiers sont présentés entre { }.

calculés; les taux d'exemption et les taux de participation sont les indicateurs de qualité les plus souvent utilisés pour les évaluations dans le domaine de l'éducation.

Les rajustements de la pondération des élèves sont créés de façon similaire aux rajustements de la pondération des écoles; les élèves qui n'ont pas participé à l'évaluation parce qu'ils étaient absents, parce qu'ils n'ont pas obtenu la permission de la faire ou parce qu'ils n'ont pas remis la feuille de réponse ou les livrets sont considérés comme des non-répondants². Ces rajustements reposent sur la théorie selon laquelle les élèves sont manquants de façon aléatoire – c'est-à-dire que les élèves qui n'ont pas participé sont similaires à ceux qui ont participé. Le rajustement pour la non-réponse de l'élève ($nr_i^{élève}$ {raj_nr_élève}), qui est calculé pour chaque classe, s'exprime de la façon suivante³ :

$$nr_i^{élève} = \frac{\# \text{ élèves participants dans la classe} + \# \text{ d'élèves non répondants dans la classe}}{\# \text{ d'élèves participants dans la classe}}$$

Pour utiliser un exemple concret, s'il y a 20 élèves dans la classe, et que 10 d'entre eux ont participé à l'évaluation, deux étaient absents et huit ont été exemptés pour diverses raisons, les 10 élèves ayant participé auraient un rajustement pour la non-réponse de 1,2 (= 12/10 = (10+2)/10).

Par conséquent, la pondération finale de l'analyse pour chaque élève ($p_i^{*élève}$ {**pondération_finale_élève**}) est obtenue de la façon suivante :

$$p_i^{*élève} = d_i^{école} * nr_{initiale_i}^{école} * nr_{rem_i}^{école} * raj_i^{classe} * nr_i^{élève},$$

ou, en utilisant les noms des variables :

$$\begin{aligned} & \text{pondération_finale_élève} \\ & = \text{pondération_conception_école} * \text{raj_nr_école} * \text{raj_sun_nr_école} * \text{raj_classe} * \\ & \quad \text{raj_nr_élève.} \end{aligned}$$

Conformément aux normes d'autres enquêtes internationales sur l'éducation, une variable dichotomique (DANS_PPCE_2016) a été créée pour indiquer si un élève a participé à l'enquête, avec :

$$dans_PPCE_2016_j = \begin{cases} 1, & \text{si l'élève } j \text{ a participé à l'évaluation} \\ 0, & \text{dans le cas contraire.} \end{cases}$$

Par conséquent, une variable de pondération finale a été créée pour les élèves participants (**pondération_finale_élèves_part**), avec :

² Voir le Tableau A – Nouveaux codes de participation, p. 20, pour la liste des codes de participation et les codes de réponse associés.

³ Le terme « classe » est défini comme une classe réelle ou un regroupement (petit) de classes pour faire une classe *de facto*.

pondération_finale_élèves_part = pondération_finale_élèves * dans_PPCE_2016.

Comme il est décrit ci-dessus, les pondérations sont dérivées à l'échelle des écoles et des élèves. Ces pondérations sont liées au volet de l'enquête qui porte sur les élèves.

Cette enquête comporte aussi deux autres volets : le *Questionnaire de l'école* et le *Questionnaire du personnel enseignant*. Puisque les données sur les écoles et le personnel enseignant n'ont pas toutes été obtenues des écoles dont les élèves participaient à l'évaluation, de nouvelles pondérations ont dû être dérivées afin de produire des estimations à l'échelle des écoles et du personnel enseignant respectivement pour les données recueillies.

Pondération de l'école (pour les données recueillies à l'échelle de l'école)

À l'échelle de l'élève, la pondération de l'école est obtenue au moyen de la formule suivante (comme il a été décrit plus haut) :

$$p_{finale_i}^{écoles} = nr_{rem_i}^{écoles} * nr_{initiale_i}^{écoles} * d_i^{écoles}.$$

La formule suivante a été utilisée, à partir de la stratification *a posteriori* appliquée pour le Québec (voir l'Annexe B) :

$$p_{finale_i}^{écoles_{pt}} = \left\{ \begin{array}{l} p_{finale_i}^{écoles} \text{ pour toutes les provinces sauf le Québec} \\ p_{finale_i}^{écoles} * STAP_{raj} \text{ pour le Québec} \end{array} \right\}.$$

À partir du fichier fourni par le CMEC, toutes les écoles pour lesquelles aucune donnée n'a été obtenue peuvent être identifiées. Le rajustement pour la non-réponse suivant est alors créé :

$$nr_{données_i}^{écoles} = \frac{\sum_i \text{écoles participantes } p_{finale_i}^{écoles_{pt}} + \sum_i \text{écoles non répondantes } p_{finale_i}^{écoles_{pt}}}{\sum_i \text{écoles participantes } P_{finale_i}^{écoles_{pt}}}.$$

La pondération finale de l'école est obtenue au moyen du calcul suivant :

$$p_{finale_i}^{données_écoles} = nr_{données_i}^{écoles} * p_{finale_i}^{écoles_{pt}}.$$

Pondération du personnel enseignant (pour les données recueillies à l'échelle du personnel enseignant)

Dans chaque école sélectionnée, l'enseignante ou l'enseignant (ou tous les membres du personnel enseignant, s'il y en a plus d'un) a été invité à remplir un questionnaire. Ainsi, la pondération de l'école et le nombre de classes sont utilisés pour calculer la pondération du personnel enseignant.

Comme il a été mentionné dans la section précédente, la pondération de l'école est obtenue au moyen de la formule suivante :

$$p_{finale_i}^{école_pt} = \begin{cases} p_{finale_i}^{école} & \text{pour toutes les provinces sauf le Québec} \\ p_{finale_i}^{école} * STAP_{raj} & \text{pour le Québec} \end{cases}$$

La pondération du personnel enseignant est donc obtenue en multipliant la pondération de l'école et le nombre de classes de 8^e année :

$$p_{finale_i}^{enseignant_pt} = p_{finale_i}^{école_pt} * raj_i^{classe}.$$

À partir du fichier « PCAP 2016 List of schools and teachers.xlsx » fourni par le CMEC, tous les membres du personnel enseignant pour lesquels aucune donnée n'a été obtenue peuvent être identifiés. Le rajustement pour la non-réponse suivant est alors créé :

$$nr_donnéese_i^{enseignants} = \frac{\sum_i \text{enseignants participants } p_{finale_i}^{enseignants} + \sum_i \text{enseignants non répondants } p_{finale_i}^{enseignants}}{\sum_i \text{enseignants participants } p_{finale_i}^{enseignants}}$$

La pondération finale du personnel enseignant est obtenue au moyen du calcul suivant :

$$p_{finale_i}^{données_enseignants} = nr_donnéese_i^{enseignants} * p_{finale_i}^{enseignant_pt}.$$

5. Pondérations bootstrap

La méthode bootstrap appartient à une famille de techniques d'estimation de la variance se fondant sur la répétition. Une discussion détaillée sur les méthodes de répétition se trouve dans Lohr (2010). De telles méthodes nécessitent l'utilisation de sous-échantillons répétés, ou répétitions, à partir des données, en calculant de nouveau l'estimation de l'enquête pondérée pour chaque répétition, et l'échantillon complet, puis en calculant ensuite la variance comme une fonction des estimations obtenues.

Pour que les analystes puissent obtenir des estimations de la variabilité qui tiennent bien compte de la conception complexe du PPCE, les pondérations bootstrap ont été créées pour les élèves participants seulement.

La méthode de Rao et Wu (1988) pour les estimations des pondérations bootstrap a été utilisée et, étant donné les fractions de sondage importantes dans beaucoup de nombreuses de strates, la variante pour les pondérations bootstrap, telles que présentées dans Beaumont et Patak

(2012) a été utilisée, le rajustement de la pondération bootstrap (a_k) étant obtenu au moyen du calcul suivant :

$$a_{h,k} = 1 - \sqrt{(1 - f_h)} + \frac{n_h}{n_h - 1} \sqrt{(1 - f_h)m_k^*}.$$

Où :

- n_h est le nombre d'écoles participantes dans la strate h ;
- f_h est le rapport entre les écoles participantes et le nombre total d'écoles dans la strate h ;
- m_k^* est le nombre de fois que l'école k a été sélectionnée pour les essais $n-1$.

Au total, 1000 pondérations bootstrap ont été créées (pbs1,...pbs1000), la pondération bootstrap d'ordre zéro (pbs0) étant la même que la pondération finale des élèves (pondération_finale_élève_part), puisque de nombreux programmes préfèrent cette structure de pondérations bootstrap.

Il importe de noter que pour les strates ayant un recensement des écoles et un recensement des classes $a_{h,k} = 1$. Voir le Tableau 5.1 pour les répartitions.

Tableau 5.1 Répartition des écoles selon les strates du recensement

Strate	Écoles		
	Nombre total	Nombre d'écoles dans l'échantillon	Nombre d'écoles participantes
Colombie-Britannique – Francophone	12	12	12
Nouveau-Brunswick – Anglophone	83	83	83
Nouveau-Brunswick – Francophone	61	61	61
Nouvelle-Écosse – Anglophone	116	116	116
Nouvelle-Écosse – Francophone	11	11	11
Île-du-Prince-Édouard – Anglophone	22	22	23
Île-du-Prince-Édouard – Francophone	3	3	3
Saskatchewan – Francophone	5	5	5

Écoles qui se sont portées volontaires	3	3	3
--	---	---	---

Les écoles qui se sont portées volontaire, c.-à-d., le remplacement de l'école sélectionnée initialement par une autre école que l'école de remplacement sélectionnée au préalable⁴, forment leur propre strate, conformément à la recommandation de Kish (1963) pour les « surprises ». Les élèves de ces écoles ont reçu une pondération de un; ils ne représentent qu'eux-mêmes, et non d'autres élèves.

Il importe de noter que, puisque le nombre d'élèves dans la classe (et l'école) au moment de la collecte est vraisemblablement différent des comptes obtenus au moment de la sélection de l'échantillon, la somme des pondérations bootstrap (pbs1...pbs1000) pour chaque répétition sera différente de la somme des pondérations finales des élèves (pbs0 = pondération_finale_élève_part).

⁴ Voir le Tableau B – Résumé des rajustements manuels des numéros d'identification des écoles, p. 21, pour de l'information supplémentaire sur les rajustements requis.

Tableau A – Nouveaux codes de participation

Participation			
Code_ Participation	Description_Code_Participation	Regroupement (Rép_élève)	Description_Rép_élève
1	Élève absent	NR	Non-répondant
2	Élève ayant participé pendant une séance prévue	Part	Participant
2A	Élève ayant participé pendant une séance prévue avec une mesure d'adaptation	Part	Participant
3	Élève ayant participé pendant une séance de rattrapage	Part	Participant
4	Élève exempté par l'école	Excl	Exclu
5	Élève exempté parce que des mesures d'adaptation appropriées ne pouvaient pas être offertes	Excl	Exclu
6	Élève n'est plus inscrit dans cette école/classe.	Quitté	A quitté de façon permanente
7	Les parents ou les élèves ne veulent plus se soumettre à l'évaluation	NR	Non-répondant
8	Élève n'est pas en 8 ^e année	Excl	Exclu
9	Élève qui fait l'école à la maison	Excl	Exclu
10	La feuille de réponse et le livret n'ont pas été remis. Seulement les données du questionnaire.	NR	Non-répondant

Tableau B – Résumé des rajustements manuels des numéros d'identification des écoles

Scénario 1 : Changements des numéros d'identification des classes au Québec

Après un travail d'enquête considérable, il a été découvert que des classes du Québec portant le numéro 00 ont été soumises à Statistique Canada, mais que, plus tard dans le cadre du processus, elles ont fait l'objet d'une conversion ne permettant pas l'utilisation du numéro de classe 00. Heureusement, le processus consistant à attribuer un nouveau numéro de classe a été découvert, et a été jugé cohérent. Essentiellement, le numéro de classe 00 est devenu le prochain numéro de classe disponible au sein de l'école. Par exemple, si l'école n'avait qu'une classe portant le numéro 00, le numéro de classe est donc devenu 01. Si l'école avait trois classes, dont les numéros étaient 00, 01 et 02, les classes portant les numéros 01 et 02 resteraient les mêmes, et la classe numéro 00 est devenue la classe 03.

Le tableau ci-dessous résume les classes sélectionnées au Québec pour lesquelles le numéro d'identification de la classe a changé et n'est plus 00. Il importe de noter que le numéro d'identification de la classe est la concaténation du numéro d'identification de l'école à quatre chiffres et du numéro de la classe à deux chiffres.

Numéro d'identification de la classe	
Soumis initialement	Après la collecte
432500	432502
232800	232801
633100	633102
633700	633706
233900	233901
234100	234101
234200	234201
634300	634305
234700	234701
635600	635601
435700	435701
635800	635801
235900	235901
236000	236001
436100	436101
436200	436201
436500	436504
437100	437101
437200	437201
637400	637401
637700	637701

Numéro d'identification de la classe	
Soumis initialement	Après la collecte
437800	437801
437900	437901
438000	438001
239100	239101
239300	239301
239500	239501
439600	439601
439700	439701
440000	440001
440300	440301
640400	640401
440500	440501
440700	440701
241006	241806
241900	241901
642200	642201
442300	442301
642500	642501

Scénario 2 : Écoles ayant reçu initialement un numéro d'identification erroné

L'explication suivante a été fournie par le CMEC : « Au début, il y avait une école comportant trois classes. Nous avons ensuite découvert que l'école était divisée en deux sections (français et anglais), et qu'elle devait donc être traitée comme deux entités distinctes. La partie 1 de l'école comportait deux classes, et la partie 2 de l'école comportait une classe. L'échantillon a été de nouveau sélectionné. Le numéro d'identification initial de la classe, qui était 220202, a dû être remplacé par 438401 ». Ce changement apporté au numéro d'identification de l'école est documenté ci-dessous :

Numéro d'identification de la classe	
Soumis initialement	Après la collecte
220202	438401

Scénario 3 : Écoles initialement exemptées, qui par la suite ont déclaré avoir des élèves visés par l'enquête

Au moment de dresser la liste, certaines écoles n'avaient pas d'élèves visés par l'enquête, ou avaient des classes sélectionnées ne comportant pas d'élèves visés par l'enquête. Même si ces écoles ont été enlevées des données pour la pondération par Statistique Canada, les données pour la pondération ont été rétablies après que des classes ont été trouvées dans la base de données des élèves participants. Ces écoles sont documentées ci-dessous :

Numéro d'identification de la classe	
Soumis initialement	Après la collecte
107301	107301
132902	132902
217702	217702

Scénario 4 : Écoles qui se sont portées volontaires – remplacement des écoles

Les écoles qui se sont portées volontaires, c.-à-d., le remplacement de l'école sélectionnée initialement par une autre école que l'école de remplacement sélectionnée au préalable, forment leur propre strate, conformément à la recommandation de Kish (1963) pour les « surprises ». Les élèves de ces écoles ont reçu une pondération de un; ils ne représentent qu'eux-mêmes, et non d'autres élèves. Ces écoles sont documentées ci-dessous :

Numéro d'identification de la classe soumis initialement	Numéro d'identification de la strate après la collecte
123202	9991_1
452301	9994_1
452301	9995_1

Références

BEAUMONT, J.F. et Z. Patak. « On the Generalized Bootstrap for Sample Surveys with Special Attention to Poisson Sampling », *International Statistical Review*, vol. 80, n° 1, 2012, p. 127-148.

KISH, L. *Survey Sampling*, John Wiley & Sons, Inc., 1963.

LOHR, S. *Sampling: Design and Analysis: 2nd Edition*, USA: Duxbury Press, 2010.

RAO, J.N.K. et C.F.J. Wu. « Resampling inference with complex survey data », *Journal American Statistical Association*, vol. 83, 1988, p. 231-241.

Rajustement pour les pondérations du Québec (stratification *a posteriori*)

Novembre 2017

Pour toutes les provinces, les pondérations finales produites ont été obtenues en fonction du plan d'échantillonnage en place et du rajustement de la non-réponse. Les variables de stratification utilisées étaient la province et la langue. Pour la province de Québec, nous avons observé que la répartition des écoles publiques et privées des répondantes et répondants était assez différente de ce que prévoyait la base de sondage pour les écoles francophones (Tableau 1). Malheureusement, cette situation rend les estimations produites moins précises.

Tableau 1 : Répartition des écoles publiques et privées : différences entre la base de sondage et l'échantillon

Strate	Échelon	Écoles publiques		Écoles privées	
		Base	Échantillon	Base	Échantillon
Québec – francophone	École	75 %	47 %	25 %	53 %
Québec – francophone	Élève	79 %	38 %	21 %	62 %
Québec – anglophone	École	75 %	77 %	25 %	23 %
Québec – anglophone	Élève	87 %	88 %	13 %	12 %

De l'information sur les écoles publiques et privées pour la population observée est disponible dans la base de sondage (Tableau 2). Cette information peut être intégrée au processus de pondération et, par la suite, rendre les estimations plus précises.

La stratification *a posteriori* est utilisée pour rajuster les pondérations de l'enquête lorsque les données auxiliaires sont sous forme de comptes. Pour la province de Québec, les comptes ont été obtenus à partir de la base de sondage.

Tableau 2 : Comptes de la base de sondage

Strate	Échelon	Comptes de populations	
		Écoles publiques	Écoles privées
Québec – francophone	École	361	119
Québec – francophone	Élève	56 739	15 170
Québec –	École	67	22

anglophone			
Québec – anglophone	Élève	6 457	968

Puisque les données auxiliaires sont disponibles aux deux échelons, la stratification *a posteriori* est utilisée pour les deux. Pour chaque échelon, le facteur de rajustement stratifié *a posteriori* est calculé pour chaque strate de second niveau (écoles publiques et écoles privées) [Tableau 3]. À l'échelon de l'école, ce facteur correspond au rapport entre le nombre d'écoles dans la strate de second niveau et le nombre estimatif d'écoles dans la strate de second niveau, qui est estimé au moyen des pondérations finales initiales des écoles. De la même façon, pour l'échelon des élèves, ce facteur correspond au rapport entre le nombre d'élèves dans la strate de second niveau et le nombre estimatif d'élèves dans la strate de second niveau, qui est estimé au moyen des pondérations finales initiales des élèves.

Les estimations du nombre d'écoles publiques et d'écoles privées (élèves) obtenues après l'application de ces facteurs de rajustement stratifiés *a posteriori* aux pondérations finales initiales correspondantes reflètent maintenant les comptes du cadre d'échantillonnage.

Tableau 3 : Facteurs de rajustement stratifiés *a posteriori* utilisés

Strate	Échelon	Facteurs de rajustement stratifiés <i>a posteriori</i> (_{raj} STP)	
		Écoles publiques	Écoles privées
Québec – francophone	École	1,5673	0,4678
Québec – francophone	Élève	2,8467	0,4738
Québec – anglophone	École	1,0372	1,1419
Québec – anglophone	Élève	1,0357	1,134

1. Pondération des écoles

À l'échelon des élèves, la pondération des écoles est obtenue au moyen de la formule suivante (voir l'Annexe A : Programme pancanadien d'évaluation – Conception de l'échantillon et pondération) :

$$p_{finale_i}^{école} = nr_{subs_i}^{école} * nr_{initiale_i}^{école} * d_i^{école}$$

Pour tenir compte de la stratification *a posteriori* précédente, une nouvelle pondération d'écoles finale est obtenue au moyen de la formule suivante :

$$p_{finale_i}^{école_SP} = \left\{ \begin{array}{l} p_{finale_i}^{école} \text{ pour toutes les provinces sauf le Québec} \\ p_{finale_i}^{école} * STP_{raj} \text{ pour le Québec} \end{array} \right\}$$

2. Pondération des élèves

À l'échelon des élèves, la pondération des écoles est obtenue au moyen de la formule suivante (voir l'Annexe A : Programme pancanadien d'évaluation – Conception de l'échantillon et pondération) :

$$p_i^{*élève} = d_i^{école} * nr_{initiale_i}^{école} * nr_{subs_i}^{école} * raj_i^{classe} * nr_i^{élève}$$

Pour tenir compte de la stratification *a posteriori* précédente, une nouvelle pondération d'élèves finale est obtenue au moyen de la formule suivante :

$$p_{finale_i}^{élève_SP} = \left\{ \begin{array}{l} p_i^{*élève} \text{ pour toutes les provinces sauf le Québec} \\ p_i^{*élève} * STP_{raj} \text{ pour le Québec} \end{array} \right\}$$

PPCE 2016 : valeurs de départ et variables de tri pour l'approche bootstrap

Les tableaux suivants présentent les valeurs de départ et les variables de tri des données avant l'application de la méthode bootstrap. Si plusieurs variables sont énumérées, les données ont été triées dans l'ordre des variables énumérées. C'est la procédure SAS *proc surveyselct* qui a été utilisée. La méthode d'échantillonnage était celle de *l'échantillonnage aléatoire sans restriction* (method=urs), avec 200 répétitions pour toutes les estimations.

Pour le rapport public du PPCE 2016

Estimation		Valeur de départ	Variable de tri
Moyenne des scores dans tous les domaines : lecture, mathématiques, sciences	Ensemble du Canada	30459584	IDELEVE
	Canada, par sexe	03951107	
	Canada, par langue du système scolaire	1024130	
	Par province	110042	
	Par province et par sexe	210342	
	Par province et par langue du système scolaire	310042	PROVINCE, LANGUE_ECOLE
Moyenne des scores dans tous les sous-domaines de la lecture	Ensemble du Canada	30459584	IDELEVE
	Canada, par sexe	03951107	
	Canada, par langue du système scolaire	10302419	
	Par province	302110	
	Par province et par sexe	210342	
	Province, par langue du système scolaire	310042	PROVINCE, LANGUE_ECOLE
Répartition en pourcentage des niveaux de compétence en lecture	Ensemble du Canada	30459584	IDELEVE
	Canada, par sexe	03951107	SEXE, IDELEVE
	Canada, par langue du système scolaire	1024130	LANGUE_ECOLE, IDELEVE
	Par province	302110	IDELEVE
	Par province et par sexe	210342	PROVINCE, SEXE, IDELEVE
	Par province et par langue	210342	PROVINCE, LANGUE_ECOLE, IDELEVE

Pour le rapport contextuel du PPCE 2016

Analyses au niveau des élèves			
Estimation		Valeur de départ	Variable de tri
Scores moyens	Première langue des élèves, par langue du système scolaire	310042	LANGUE_ECOLE
	Programme de langue seconde des élèves, par langue du système scolaire	310042	LANGUE_ECOLE
	Autres variables relatives aux élèves	03951107	IDELEVE
Moyenne des indices	Ensemble du Canada	30459584	IDELEVE
	Canada, par sexe, ou Canada, par langue	1024130	IDELEVE
	Par province et par sexe	110042	IDELEVE
	Par province et par langue	310042	PROVINCE, LANGUE_ECOLE
Moyenne par quartile des indices	Ensemble	03951107	IDELEVE

Analyses au niveau du personnel enseignant			
Estimation		Valeur de départ	Variable de tri
Moyenne des élèves (moyenne de moyenne)	Toutes les variables relatives au personnel enseignant	03951107	IDENSEIGNANT
Moyenne des indices	Ensemble du Canada	30459584	IDENSEIGNANT
	Canada, par langue du système scolaire	1024130	
	Par province	110042	
	Par province et par langue du système scolaire	310042	PROVINCE, LANGUE_ECOLE
Moyenne par quartile des indices	Ensemble	03951107	IDENSEIGNANT