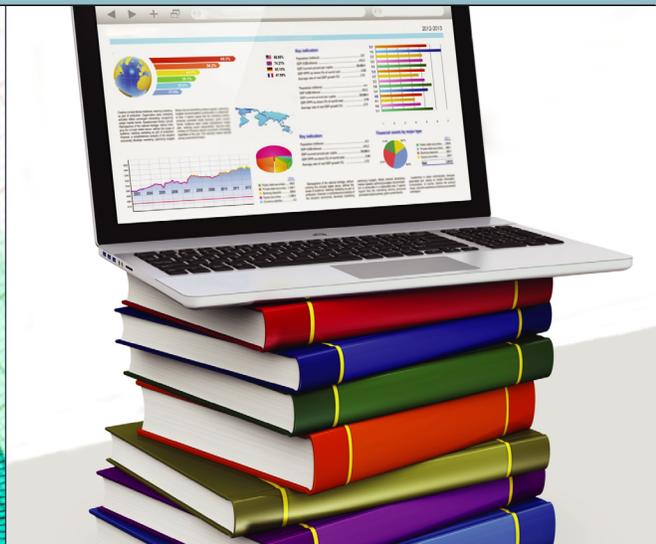
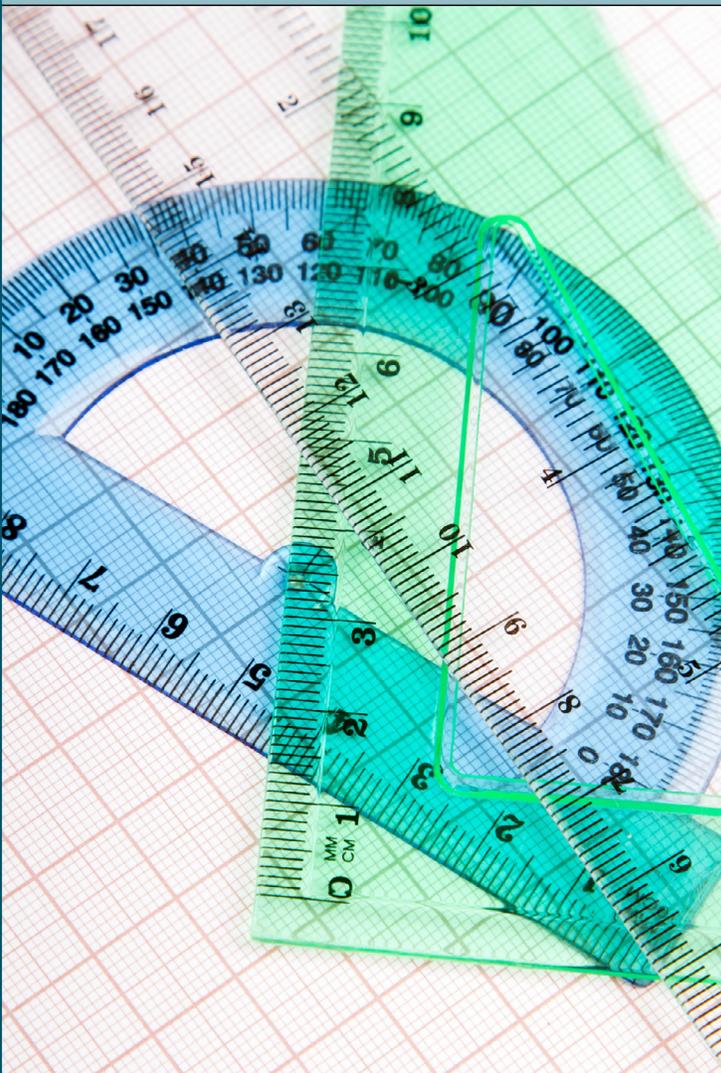


PPCE 2019

Rapport technique



Programme pancanadien d'évaluation

PPCE 2019

Rapport technique

Auteurs et auteurs

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

M. Koffi Houme, Conseil des ministres de l'Éducation (Canada)

M. Yitian Tao, Conseil des ministres de l'Éducation (Canada)

M. Gulam Khan, Conseil des ministres de l'Éducation (Canada)

M^{me} Allison Chapman-Chin, Conseil des ministres de l'Éducation (Canada)



cmec

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 provinciaux et territoriaux 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 les 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 territoires. 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 secondaire.

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

© 2022 Conseil des ministres de l'Éducation (Canada)

This report is also available in English.

TABLE DES MATIÈRES

| | |
|---|-----------|
| Chapitre 1. Vue d'ensemble du Programme pancanadien d'évaluation | 1 |
| Contexte | 1 |
| Évaluation pancanadienne | 1 |
| Participation | 2 |
| Durée de l'évaluation..... | 2 |
| Le PPCE dans les deux langues officielles | 3 |
| | |
| Chapitre 2. Conception et élaboration de l'évaluation | 4 |
| Cycle d'évaluation du PPCE..... | 4 |
| Présentation des résultats du PPCE au fil du temps | 5 |
| Mise à jour du cadre d'évaluation | 5 |
| Transition au PPCE en ligne | 5 |
| Conception de l'évaluation | 6 |
| Groupes de travail | 7 |
| Mise à jour du cadre d'évaluation en mathématiques | 7 |
| Élaboration des items | 8 |
| Examen des items | 8 |
| Vérification des items d'évaluation et des guides de codage | 9 |
| Révisions linguistiques et typographiques | 9 |
| Révisions techniques | 9 |
| Révisions psychométriques | 10 |
| Approbation des items par les provinces | 11 |
| | |
| Chapitre 3. Élaboration des questionnaires contextuels..... | 12 |
| Mise à jour du cadre des questionnaires | 12 |
| Identité sexuelle | 13 |
| Différences entre les sexes | 13 |
| Confidentialité | 14 |
| | |
| Chapitre 4. Procédures d'échantillonnage | 15 |
| Plan d'échantillonnage | 15 |
| Base d'échantillonnage | 16 |
| Première étape de la base d'échantillonnage – liste des écoles admissibles | 16 |
| Deuxième étape de la base d'échantillonnage – liste des classes admissibles | 16 |
| Stratification | 17 |
| Taille des échantillons..... | 17 |
| Première étape de l'échantillonnage – échantillonnage des écoles | 18 |
| Bases de données sur les écoles | 18 |
| Sélection des écoles | 18 |
| Exclusion d'écoles..... | 18 |

| | |
|--|-----------|
| Deuxième étape de l'échantillonnage – échantillonnage des élèves | 19 |
| Pondération de l'échantillon..... | 21 |
| Chapitre 5. Mise à l'essai..... | 22 |
| Examen des items | 22 |
| Livrets d'évaluation..... | 22 |
| Séance de codage | 22 |
| Saisie des données | 23 |
| Analyse des données | 23 |
| Sélection des items | 24 |
| Révision du cadre d'évaluation | 24 |
| Chapitre 6. Étude principale..... | 25 |
| Livrets d'évaluation..... | 25 |
| Réexamen du matériel d'évaluation | 25 |
| Évaluation sur papier..... | 25 |
| Évaluation en ligne | 25 |
| Documents administratifs..... | 26 |
| Lettre envoyée aux parents/tutrices ou tuteurs des élèves | 26 |
| Procédures administratives | 26 |
| Lieu de l'évaluation..... | 27 |
| Déroulement de l'évaluation | 27 |
| Élèves ayant des besoins spéciaux..... | 27 |
| Questionnaires pour la direction de l'école et pour le personnel enseignant..... | 28 |
| Participation des élèves à l'évaluation et exemptions | 29 |
| Organisation d'une séance de rattrapage..... | 30 |
| Renvoi du matériel d'évaluation | 31 |
| Séance de codage | 31 |
| Codage | 31 |
| <i>Manuel de la codeuse ou du codeur.....</i> | 31 |
| <i>Guide de codage.....</i> | 31 |
| Chefs de codage | 32 |
| Chefs de table..... | 32 |
| Formation des codeuses et codeurs..... | 32 |
| Fidélité du codage | 33 |
| Examens de fidélité | 33 |
| Fiabilité entre membres du groupe de codage (codage multiple) | 34 |
| Fiabilité des tendances..... | 34 |
| Rapports et commentaires | 34 |
| <i>Rapport de la coordonnatrice ou du coordonnateur de la province.....</i> | 35 |
| Chapitre 7. Évaluation en ligne..... | 37 |
| Équivalence et examen des items..... | 37 |
| Configuration informatique requise et déploiement du test..... | 37 |
| Saisie des données et codage des réponses des élèves | 38 |

| | |
|---|-----------|
| Chapitre 8. Étude de la transition à une évaluation en ligne | 40 |
| Aperçu de l'étude sur le mode d'administration dans le cadre de la mise à l'essai | 40 |
| Aperçu de l'étude sur le mode d'administration dans le cadre de l'étude principale..... | 42 |
| Chapitre 9. Établir une norme de rendement | 43 |
| Séances d'établissement des normes..... | 43 |
| Sélection d'un groupe de spécialistes..... | 43 |
| Versions préliminaires des descripteurs des niveaux de rendement..... | 44 |
| Sécurité des documents | 44 |
| Méthode du signet | 44 |
| Procédure d'établissement des normes | 45 |
| Descripteurs des niveaux de rendement..... | 48 |
| Chapitre 10. Traitement des données du PPCE | 49 |
| Nettoyage des données | 49 |
| Recodage général | 49 |
| Examen des données d'échantillonnage | 50 |
| Dernier examen des données et préparation de la base de données | 50 |
| Chapitre 11. Analyse des données sur le rendement des élèves | 51 |
| Analyse préliminaire..... | 51 |
| Filtrage des données | 51 |
| Recodage des items..... | 52 |
| Données manquantes | 53 |
| Analyse des items..... | 54 |
| Analyse des items dans le cadre de la théorie classique des tests..... | 54 |
| Analyse de la TRI..... | 57 |
| Évaluation de la dimensionnalité du PPCE 2019 | 57 |
| Calibrage des items et évaluation du caractère convenable des modèles de la TRI .. | 57 |
| Invariance du PPCE 2019 | 58 |
| Fonctionnement du test | 63 |
| Liens et équivalences avec les évaluations précédentes pour les domaines secondaires | 63 |
| Production des scores pour le rendement et scores à l'échelle | 64 |
| Estimations des erreurs-types | 64 |
| Présentation des résultats de l'évaluation du PPCE 2019..... | 65 |
| Chapitre 12. Analyse des données du questionnaire..... | 66 |
| Analyse préliminaire | 66 |
| Filtrage des données | 66 |
| Recodage des items..... | 67 |
| Données manquantes | 67 |

| | |
|--|-----------|
| Statistiques descriptives | 68 |
| Analyse corrélacionnelle | 68 |
| Analyse en composantes principales | 68 |
| Analyses des items et des indices..... | 71 |
| Analyse de comparaison des groupes | 71 |
| Chapitre 13. Ensembles de données du PPCE | 73 |
| Description des ensembles de données | 73 |
| Ensemble de données des élèves | 73 |
| Ensemble de données du personnel enseignant | 73 |
| Ensemble de données des écoles | 74 |
| Ensemble de données fusionné – élèves/personnel enseignant/écoles | 74 |
| Consultation de l'ensemble de données pour les recherches | 75 |
| Conditions et modalités..... | 75 |
| Information pour les chercheuses et chercheurs | 76 |
| Bibliographie..... | 78 |
| ANNEXE A : Plan d'échantillonnage des écoles et des classes..... | 82 |
| 1. Introduction | 82 |
| 2. Plan d'échantillonnage..... | 82 |
| 2.1 Populations cibles et populations de l'enquête..... | 83 |
| 2.2 Base d'échantillonnage..... | 83 |
| 2.2.1 Première étape de la base d'échantillonnage (liste des écoles à l'intérieur du champ)..... | 84 |
| 2.2.2 Deuxième étape de la base d'échantillonnage (liste des classes à l'intérieur du champ)..... | 85 |
| 2.3 Stratification | 86 |
| 2.4 Attribution de la taille de l'échantillon | 86 |
| 2.5 Sélection de l'échantillon des écoles et des classes | 88 |
| 2.5.1 Chevauchement minimal avec la TEIMS 2019..... | 89 |
| 3. Pondérations d'échantillonnage | 89 |
| 3.1 Rajustements de la pondération d'échantillonnage pour la non-réponse | 90 |
| 3.2 Pondérations et rajustements à l'échelle de la classe | 92 |
| 4. Pondérations bootstrap | 93 |

LISTE DES FIGURES

| | | |
|--------------------|--|----|
| FIGURE 8.1 | Conception des livrets pour l'étude sur le mode d'administration de la mise à l'essai | 41 |
| FIGURE 11.1 | Détection du FDI selon le mode : courbes caractéristiques du test pour les items de mathématiques..... | 61 |
| FIGURE 11.2 | Détection du FDI selon le sexe : courbes caractéristiques du test pour les items de mathématiques..... | 62 |
| FIGURE 11.3 | Détection du FDI selon la langue : courbes caractéristiques du test pour les items de mathématiques..... | 62 |

LISTE DES TABLEAUX

| | | |
|---------------------|--|----|
| TABLEAU 2.1 | Calendrier des évaluations du PPCE | 4 |
| TABLEAU 2.2 | Nombre de blocs, de scénarios et d'items par domaine et par livret | 7 |
| TABLEAU 2.3 | Répartition des items selon le type et le domaine d'évaluation | 7 |
| TABLEAU 4.1 | Résumé des critères utilisés pour échantillonner les élèves, selon les types de strates..... | 20 |
| TABLEAU 7.1 | Configuration informatique requise pour le PPCE en ligne | 37 |
| TABLEAU 9.1 | Répartition des élèves par niveau de rendement en mathématiques | 47 |
| TABLEAU 11.1 | Classification du FDI selon Educational Testing Service (ETS)..... | 60 |
| TABLEAU 11.2 | Fourchettes des mesures de l'ampleur de l'effet pour les items avec FDI selon le mode..... | 60 |
| TABLEAU 11.3 | Fourchettes des mesures de l'ampleur de l'effet pour les items avec FDI selon la langue et le sexe | 61 |
| TABLEAU A.1 | Variables importantes pour la première étape de la sélection de l'échantillon | 84 |
| TABLEAU A.2 | Exclusions des écoles par type | 84 |
| TABLEAU A.3 | Couverture de la base du PPCE | 85 |
| TABLEAU A.4 | Variables importantes pour la deuxième étape de la sélection de l'échantillon | 85 |
| TABLEAU A.5 | Attribution de la taille de l'échantillon des écoles – PPCE 2019..... | 87 |
| TABLEAU A.6 | Répartition des écoles selon les strates du recensement | 94 |
| TABLEAU A.7 | Codes de participation | 94 |

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, 2018 et 2022 (avec plus de 80 pays et économies en 2018), au Programme international de recherche en lecture scolaire (PIRLS) de 2011, 2016 et 2021 (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 et des 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 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 (et se déroulera encore en 2022) 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 Programme pancanadien d'évaluation (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 province à l'autre et d'un territoire 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 secondaire¹ 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 les évaluations ne sont pas conçues 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, en 2016 et en 2019, 10 provinces ont participé au PPCE. Les Territoires du Nord-Ouest ont, par le passé, participé au PIRS.

Durée de l'évaluation

Au PPCE 2019, les élèves se sont vu accorder 90 minutes pour répondre aux items de l'évaluation. 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 au test cognitif, 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.

¹ Le PPCE vise les élèves de 2^e secondaire au Québec et de 8^e année dans le reste du Canada.

Le PPCE dans les deux langues officielles

Les résultats du Canada francophone sont ceux des élèves qui fréquentent les systèmes d'éducation de langue française de leur province respective. De la même manière, les résultats du Canada anglophone sont ceux des élèves qui fréquentent les systèmes d'éducation de langue anglaise de leur province respective. Dans les systèmes scolaires anglophones, même si les élèves des programmes d'immersion française pouvaient, à la discrétion de leur école, passer l'évaluation du PPCE en anglais ou en français, leurs résultats sont présentés avec ceux de la cohorte de langue anglaise. La liste des termes scientifiques et mathématiques courants en anglais et en français a été fournie comme ressource à ces élèves.



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 minoritaire² 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 des échantillons, les résultats des systèmes scolaires francophones de l'Île-du-Prince-Édouard et de Terre-Neuve-et-Labrador ont été exclus; en revanche, ils ont été inclus dans le calcul des moyennes et des totaux des provinces et de l'ensemble du Canada.

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 2019, la cinquième évaluation était axée sur les mathématiques, comme cela avait été le cas en 2010 – la lecture et les sciences constituant les domaines secondaires.

TABEAU 2.1 Calendrier des évaluations du PPCE

| Domaine | Cycle 1 | | | Cycle 2 | | |
|------------|----------------|----------------|----------------|----------------|----------------|-----------------------------|
| | Printemps 2007 | Printemps 2010 | Printemps 2013 | Printemps 2016 | Printemps 2019 | Printemps 2023 ³ |
| Principal | Lecture | Mathématiques | Sciences | Lecture | Mathématiques | Sciences |
| Secondaire | Mathématiques | Sciences | Lecture | Mathématiques | Sciences | Lecture |
| Secondaire | Sciences | Lecture | Mathématiques | Sciences | Lecture | Mathématiques |

² En ce qui concerne la ventilation des deux langues officielles du Canada, l'anglais est la langue majoritaire en dehors du Québec – dans l'ensemble du pays, 64 p. 100 des Canadiennes et Canadiens indiquent parler anglais la plupart du temps à la maison. Au Québec, le français est la langue majoritaire – 79 p. 100 de la population québécoise indique parler français la plupart du temps à la maison (Statistique Canada, 2020).

³ Le déploiement du PPCE 2022 a été reporté à 2023 en raison de préoccupations liées à la santé soulevées par la pandémie mondiale et pour limiter le chevauchement avec le PISA, qui a été retardé d'une année, soit de 2021 à 2022.

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, 2016 et 2019, 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 2019 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 et divers territoires à une étape comparable de leur scolarité. Le *Cadre d'évaluation du PPCE 2019* (CMEC, 2020) présente les bases théoriques, les principes de conception et les descripteurs de rendement utilisés dans l'élaboration des items des trois domaines pour le deuxième cycle du PPCE (2016-2023). Le chapitre 2 décrit le domaine principal, soit les mathématiques, tandis que les chapitres 3 et 4 décrivent les domaines secondaires, soit les sciences et la lecture, respectivement.

Pour l'année 2019, le cadre d'évaluation des mathématiques a été mis à jour pour mieux refléter les programmes d'études et les normes à l'échelle du Canada. Le cadre d'évaluation de la lecture avait été mis à jour pour le PPCE 2016, quand la lecture a été le domaine principal pour la deuxième fois, et celui 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 pour ces deux domaines. Les spécialistes provinciaux ont discuté des mises à jour et y ont apporté des révisions.

Transition au PPCE en ligne

En 2019, pour la première fois, l'évaluation du PPCE a été réalisée en ligne. Pour tenir compte des effets du passage du test sur papier à un test en ligne et pour pouvoir établir des liens entre les données des deux versions, ou modes, une étude sur les effets du mode a été réalisée sur un échantillon réduit d'élèves, qui ont passé le test cognitif et répondu au questionnaire sur papier. Après avoir comparé les résultats de l'évaluation sur papier et de l'évaluation en ligne, le Groupe consultatif technique du PPCE a conclu qu'ils étaient comparables, tant pour ce qui est de l'évaluation de 2019 que des comparaisons au fil du temps. Les questionnaires du personnel enseignant et de l'école ont tous été remplis en ligne en 2019.

Conception de l'évaluation

L'évaluation du PPCE couvre trois domaines d'évaluation : les mathématiques, la lecture et les sciences. Tous les domaines font l'objet d'une rotation. Cette rotation s'échelonne habituellement sur une période de neuf ans, qui débute par le domaine principal, avec un cadre d'évaluation nouveau ou mis à jour à la lumière des meilleures connaissances sur l'évaluation de cette matière. La rotation se poursuit avec les deux évaluations suivantes du PPCE, dans lesquelles cette même matière est un domaine secondaire. La rotation se termine avec cette matière de nouveau comme domaine principal, avec une autre mise à jour du cadre d'évaluation. Lors de l'évaluation de 2019, le domaine principal était les mathématiques, tandis que la lecture et les sciences étaient les domaines secondaires. Le domaine principal change à chaque édition de l'évaluation; cela signifie que les mathématiques seront un domaine secondaire et que les sciences seront le domaine principal du PPCE en 2023 (un décalage d'un an par rapport à la rotation habituelle, en raison de la pandémie mondiale).

Pour que les tendances puissent être mesurées sur de plus longues périodes, chaque fois que le cadre d'évaluation du domaine principal est mis à jour – soit au début de chaque rotation pour ce domaine –, de nouveaux items sont élaborés en fonction de l'évolution du concept hypothétique. Pour le PPCE 2019, le cadre révisé d'évaluation des mathématiques et l'introduction d'items pour une évaluation assistée par ordinateur ont élargi le concept hypothétique⁴ au-delà de celui mesuré en 2010, soit la dernière fois où les mathématiques étaient le domaine principal. Par conséquent, l'échelle des mathématiques du PPCE 2019 doit correspondre au cadre mis à jour, tout en étant liée à l'échelle du cadre précédent par l'ensemble d'items d'ancrage existants.

Pour l'évaluation du PPCE, les unités d'évaluation en mathématiques ont été regroupées en huit groupes ou « blocs » répartis entre quatre versions, ou livrets, de sorte que chaque livret contenait généralement deux blocs d'items en mathématiques, un bloc d'items en lecture et un bloc d'items en sciences. 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 mathématiques; 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, pour permettre de faire des mesures comparatives du rendement des élèves d'un livret à l'autre. Après le test cognitif, les élèves ont rempli le *Questionnaire de l'élève*.

Le tableau 2.2 montre la répartition des blocs, des contextes (ou scénarios ou passages) et des items pour les mathématiques, la lecture 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.

⁴ Le « concept hypothétique » est la traduction du mot anglais « construct », suggérée par Laveault et Grégoire (2002, 2014). Certains logiciels de traduction, p. ex., *Logiterm*, indiquent que le mot est adopté en français comme tel ou le traduisent par « concept ». Certains chercheurs et auteurs ont utilisé une différente terminologie. Par exemple, Bertrand et Blais (2004) utilisent le mot « construit ». La traduction retenue dans le présent document a l'avantage d'une précision supplémentaire : il ne s'agit pas seulement d'un concept (ou construit) en général, mais ce concept est hypothétique, ne pouvant être observé ou directement mesuré.

TABLEAU 2.2 Nombre de blocs, de scénarios et d'items par domaine et par livret

| | Mathématiques | | | Lecture | | | Sciences | | |
|-----------------|---------------|-----------|-------|---------|----------|-------|----------|-----------|-------|
| | Blocs | Scénarios | Items | Blocs | Passages | Items | Blocs | Scénarios | Items |
| Livret 1 | 2 | 7 | 22 | 1 | 2 | 10 | 1 | 3 | 9 |
| Livret 2 | 2 | 7 | 19 | 1 | 3 | 12 | 1 | 3 | 8 |
| Livret 3 | 2 | 7 | 15 | 1 | 3 | 9 | 1 | 3 | 8 |
| Livret 4 | 2 | 6 | 17 | 1 | 3 | 9 | 1 | 3 | 8 |

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 |
| Mathématiques | 17 | 5 | 14 | 5 | 13 | 6 | 14 | 6 |
| Lecture | 10 | 2 | 12 | 2 | 7 | 2 | 7 | 2 |
| Sciences | 7 | 2 | 7 | 3 | 8 | 2 | 11 | 1 |
| Total | 34 | 9 | 33 | 10 | 28 | 10 | 32 | 9 |

* Réponse choisie

** Réponse construite

Groupes de travail

Le domaine principal du PPCE 2019 était les mathématiques. 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 des mathématiques. Les groupes de travail se composaient de spécialistes des programmes d'études des mathématiques, 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 codage 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 mathématiques

Des spécialistes des mathématiques ont examiné et mis à jour le cadre d'évaluation pour qu'il reflète la pensée actuelle dans le domaine de la recherche sur la culture mathématique et la numératie, y compris l'évaluation en mathématiques.

Le groupe de travail chargé de l'examen du cadre d'évaluation en mathématiques a réuni des spécialistes des programmes d'études des mathématiques et de l'évaluation, qui se sont penchés sur ledit cadre et y ont apporté des modifications supplémentaires pour refléter les changements dans les programmes d'études des mathématiques en vigueur dans les provinces de l'ensemble du Canada.

Élaboration des items

Les documents devant servir à guider toutes les étapes du processus d'élaboration des items ont été préparés pour les réunions des responsables de l'élaboration des tests, qui ont eu lieu à Ottawa en juillet 2017 et à Toronto en août 2017.

Le processus d'initiation comprenait une vue d'ensemble du cadre d'évaluation en mathématiques, 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 qui correspondent 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 mathématiques était bien couvert.

À la fin des séances d'élaboration des items, le groupe de travail a réexaminé et révisé le cadre d'évaluation en mathématiques 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 mathématiques des évaluations antérieures pour s'assurer qu'ils représentaient bien le cadre d'évaluation et avaient bien la bonne classification.

Les items ont été élaborés à la fois en français et en anglais et, une fois terminés, ont été traduits et révisés au Secrétariat du CMEC.

Examen des items

Trois livrets ont été conçus de façon à contenir à peu près le même nombre d'items sur chacun des quatre sous-domaines des mathématiques (nombres et opérations, géométrie et mesure, régularités et relations ainsi que gestion des données et probabilités) et le même nombre d'items de chaque niveau cognitif et de chaque type (réponse construite et réponse choisie). Ils ont été présentés à des spécialistes provinciaux des mathématiques. Ceux-ci ont examiné les items pour en 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.

À la fin de cet examen, 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 mathématiques conformément au cadre d'évaluation en mathématiques;
- vérification des guides de codage pour les items à réponse construite;
- révisions linguistiques et stylistiques, révisions techniques et révisions psychométriques.

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 la même dans les livrets en ligne et les livrets papier, que chaque item commençait par un énoncé suivi d'une question, que l'espace fourni 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 que les descripteurs corrects avaient bien été attribués et que leur exactitude avait été vérifiée, 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 mathématiques, en lecture et en sciences 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. Lorsque les réponses possibles étaient constituées d'un texte, les leurres étaient disposés en ordre croissant de longueur, 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. 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 mathématiques 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 et de 2016 pour la lecture, l'évaluation de 2010 pour les mathématiques et l'évaluation de 2013 pour les sciences), afin de garantir la cohérence des codages 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. 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 codage et s'assurer que les codes définis étaient clairs et précis. Cette étape était très importante, parce que, lors de la séance de codage des items dans les

trois matières, les codeuses et codeurs 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.

Approbation des items par les provinces

Avant d'inclure les items dans la mise à l'essai et l'étude principale, les items choisis devaient être approuvés par les provinces. Le CMEC a produit trois livrets pour la mise à l'essai et quatre livrets pour l'étude principale, en français et en anglais, à la fois sur papier et en ligne. Il 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 2018 et l'étude principale en 2019.

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 se servir de l'information fournie par ces questionnaires pour déterminer quels facteurs influent sur l'apprentissage. Le contenu des questionnaires contextuels varie selon laquelle des trois matières est évaluée comme domaine principal du PPCE. Comme le domaine principal du PPCE 2019 était les mathématiques, certaines questions contextuelles du *Questionnaire de l'élève* portaient notamment sur des facteurs qui, selon des études antérieures, sont en corrélation avec le rendement en mathématiques. Entre autres exemples de ces facteurs corrélés, il y a le niveau de scolarité des parents, la langue parlée à la maison et le nombre de livres à la maison.

Les questionnaires contextuels remplis par le personnel enseignant portent sur les conditions d'enseignement et d'apprentissage, y compris les attentes des enseignantes et enseignants au chapitre des devoirs, leurs méthodes d'évaluation, leurs domaines de spécialisation et leur nombre d'années d'expérience en enseignement. Le *Questionnaire de l'école*, rempli par la direction, est la principale source d'informations sur diverses dimensions de l'école : structure et organisation de l'école; climat scolaire; politiques et pratiques de l'école; programmes et enseignement.

Les questionnaires du PPCE sont publiés sur le site Web du CMEC à l'adresse https://cmec.ca/697/PCAP_2019.html. L'ensemble de données du PPCE est accessible sur demande.

Le chapitre 5 du *Cadre d'évaluation du PPCE 2019* (CMEC, 2020) 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 de l'apprentissage des mathématiques.

Mise à jour du cadre des questionnaires

Il a été demandé aux spécialistes des mathématiques d'examiner et de réviser, indépendamment les uns des autres, le cadre des questionnaires utilisé pour le PPCE 2010, l'année précédente où les mathématiques étaient le domaine principal de l'évaluation. Il leur a été demandé de cerner les importants aspects du contexte dans lequel les élèves apprennent les mathématiques, ainsi que de cerner les domaines qui pourraient être améliorés ou les sujets qui pourraient être ajoutés selon la documentation universitaire sur les mathématiques, qui fournit un contexte à la manière dont les élèves apprennent les mathématiques 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 des mathématiques – soit le domaine principal – et qui pourraient fournir d'importants renseignements contextuels aux provinces et aux territoires. Ces questionnaires axés sur les mathématiques 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 2019 comprenait trois questionnaires : un questionnaire pour les élèves participant à l'évaluation, un questionnaire pour leurs enseignantes et enseignants de mathématiques 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 mathématiques 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 socioéconomique 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 les mathématiques, 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 sur les mathématiques.

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 et du PPCE 2019, la question du genre a été élargie pour inclure deux choix supplémentaires pour les personnes interrogées, comme le montre l'encadré ci-dessous.

| Comment t'identifies-tu? |
|---|
| <input type="radio"/> Garçon/Homme |
| <input type="radio"/> Fille/Femme |
| <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 mathématiques en faveur des élèves de sexe masculin 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 les mathématiques hors de l'école.

Les différences entre les sexes persistent également en lecture, en faveur des filles cette fois-ci. Même si, pour ce qui est 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 dans les trois domaines.

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.

Au printemps 2019, la cinquième évaluation du PPCE a été réalisée. Trois domaines ont été évalués : les mathématiques, la lecture et les sciences – les mathématiques 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 les mathématiques. Un seul niveau scolaire (8^e année/2^e secondaire) a été évalué. Dix-huit populations⁵ 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 2019, comme lors de l'évaluation de 2016, le CMEC a retenu les services de Statistique Canada et a travaillé en étroite collaboration avec celui-ci sur 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 2019; 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 cible⁶ facilite le processus d'échantillonnage et évite les ambiguïtés. Le tableau A.3 de l'Annexe A montre la population représentée dans le plan d'échantillonnage du PPCE. Pour valider l'exactitude de la base d'échantillonnage, 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 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 2018.

⁵ Le mot « population » désigne l'ensemble des élèves de 8^e année/2^e secondaire admissibles dans une province ou un groupe linguistique donné.

⁶ La « population cible » fait référence aux écoles admissibles à la sélection après l'exclusion de celles ne répondant pas aux critères retenus par le CMEC ou les provinces et les territoires concernés. La « population globale », pour sa part, est formée de toutes les écoles accueillant des élèves de 8^e année/2^e secondaire.

Base d'échantillonnage

Première étape de la base d'échantillonnage – liste des écoles admissibles⁷

Comme pour les évaluations du PIRS et les précédents cycles du PPCE, une procédure d'échantillonnage à deux étapes a été utilisée pour le PPCE 2019 : dans un premier temps, la sélection d'écoles participantes (première étape de la base d'échantillonnage – 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 (deuxième étape de la base d'échantillonnage – 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 critères⁸. 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.

Dans la préparation de la base d'échantillonnage, 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'à la première étape de la base d'échantillonnage, les écoles exemptées ont été exclues : elles n'ont donc pas été prises en compte durant le processus d'échantillonnage.

Deuxième étape de la base d'échantillonnage – liste des classes admissibles

À la deuxième étape de la base d'échantillonnage, la liste des classes admissibles a été obtenue. Cette base d'échantillonnage 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 :

- incapacités fonctionnelles;
- incapacités intellectuelles ou troubles socioaffectifs;

⁷ Le mot « admissible » est la traduction de l'expression anglaise « *in-scope* ». Également traduite par « à l'intérieur du champ », l'expression est utilisée relativement aux unités ou sujets de la population cible qui peuvent être inclus dans l'échantillon. Elle distingue ces unités de celles qui sont **inadmissibles** (« *out-of-scope* » *units*).

⁸ Dans le cas du recensement des élèves, il n'y a pas d'inférence statistique et les marges d'erreur n'ont habituellement pas besoin d'être compilées.

- 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 de la province 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 A.5 de l'Annexe A présente la liste de strates et indique si les écoles ont été échantillonnées à la première étape ou si les recensements ont été utilisés au niveau des écoles.

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 qui peuvent être acceptés 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. La formule suivante a été utilisée pour déterminer la taille d'un échantillon par rapport au calcul des distributions de fréquences.

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

où :

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 qui peut être acceptée

Première étape de l'échantillonnage – 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 de ses 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.

Deuxième étape de l'échantillonnage – 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 2019. 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 4.1 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.

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.

Examen des items

Deux spécialistes de contenu mathématique ont examiné tous les items de mathématiques et vérifié les clés et les guides de codage ainsi que la classification des items selon les sous-domaines du test, conformément au cadre d'évaluation en mathématiques. Ils ont 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.). Les items sélectionnés pour la mise à l'essai couvraient bien les différents sous-domaines des mathématiques et différents niveaux de difficulté.

Livrets d'évaluation

Après le dernier examen, les nouveaux blocs d'items et les blocs d'items d'ancrage existants ont été réunis en différentes combinaisons de blocs numériques, utilisés à la fois pour le test en ligne et le test sur papier du PPCE. Trois versions du test, ou livrets d'évaluation, ont été préparées pour la mise à l'essai. Chaque livret était conforme aux spécifications figurant dans le *Cadre d'évaluation du PPCE 2019* et contenait environ 40 items de mathématiques, ainsi que le *Questionnaire de l'élève*. Les livrets de la mise à l'essai ont été envoyés aux provinces pour y être examinés par des spécialistes des programmes d'études et de l'évaluation. Les provinces ont de plus été invitées à vérifier que les items rédigés en français et en anglais étaient équivalents. Une fois les modifications recommandées apportées, les livrets ont été ajoutés au Système central de codage (SCC) en ligne.

Les élèves avaient 90 minutes pour remplir les items de mathématiques du livret et 30 minutes pour remplir le *Questionnaire de l'élève*. Les questionnaires du personnel enseignant et de l'école ont été préparés sous la forme de livrets séparés.

Séance de codage

La séance de codage s'est déroulée sur cinq jours à Ottawa, en juillet 2019. Approximativement 2 000 livrets ont été codés, dont environ 1 000 livrets en français et 1 000 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. Environ les deux tiers des codeuses et codeurs avaient été nommés par les ministères provinciaux responsables de l'éducation.

Le processus de codage a employé des tests de fiabilité croisés, deux fois par jour, pour s'assurer que les codeuses et codeurs évaluaient les items de la même façon et selon les codes attribués par les spécialistes.

Le degré d'uniformité entre les codeuses et codeurs 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 codeuses et codeurs ont réexaminé les documents de formation et procédé ensuite à un nouveau codage 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) à l'aide de boutons radio ou de menus déroulant dans l'évaluation numérique ou sur un feuillet de réponses détachable dans l'évaluation sur papier. Ils ont écrit leurs réponses aux items à réponse construite dans l'espace prévu dans l'évaluation numérique ou directement dans les livrets d'évaluation. 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.

Toutes les réponses des élèves ont été codées à l'aide du système de codage en ligne. Les réponses sur papier ont été scannées et saisies dans le système. Lors de la séance de codage pour la mise à l'essai, les codeuses et codeurs n'ont codé que les items à réponse construite, en choisissant le code pour chaque item qui correspondait le mieux à la réponse de l'élève. Les items ont été distribués de façon aléatoire aux codeuses et codeurs. Les questionnaires du personnel enseignant et de l'école ont été remplis en ligne. Les données recueillies grâce à l'évaluation et à tous les questionnaires 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 mathématiques ont été analysées par un spécialiste de la psychométrie du CMEC. 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).

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 la langue et le sexe.

Sélection des items

Après la mise à l'essai des items, le groupe de travail sur la sélection des items, composé de représentantes et représentants de six 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 codage des items de la mise à l'essai (à partir du questionnaire qui leur avait été donné à la fin de la séance de codage). 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 : mathématiques, lecture et sciences. Tous les items de mathématiques, de lecture 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 2007 et 2016, mathématiques pour le PPCE 2010 et sciences pour le PPCE 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 sur l'élaboration des questionnaires, qui avait préparé les trois questionnaires de la mise à l'essai, a été convoqué de nouveau. Il a 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 2019* 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.

L'évaluation du PPCE a eu lieu au printemps 2019, avec les mathématiques comme domaine principal. (Les domaines secondaires étaient la lecture et les sciences.) Près de 30 000 élèves, sélectionnés de façon aléatoire dans plus de 1 500 écoles dans 10 provinces du Canada, ont participé à l'évaluation en français ou 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. La majorité des élèves ont rempli l'évaluation en ligne; un échantillon d'élèves a toutefois rempli le test sur papier dans le cadre d'une étude sur le mode d'administration.

Livrets d'évaluation

Chaque livret d'évaluation contenait deux blocs d'items en mathématiques, un bloc d'items en lecture 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.

Évaluation sur papier

Les livrets d'évaluation ont fait l'objet d'une vérification, qui visait à s'assurer que toutes les modifications de contenu apportées par les groupes de travail et le personnel du Secrétariat du CMEC avaient été intégrées aux nouvelles versions et que les versions papier et les versions numériques du test étaient identiques. 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.

Évaluation en ligne

Le chapitre 7 décrit en détail les livrets d'évaluation et le déroulement de l'évaluation en ligne.

Documents administratifs

Les documents administratifs ont été imprimés et envoyés à chaque école participante. Chaque coordonnatrice ou coordonnateur de l'école devait s'assurer qu'elle ou il disposait des documents destinés à leur école (y compris les livrets d'évaluation destinés aux écoles choisies pour le test sur papier). Tout document manquant devait être immédiatement signalé au CMEC, afin que celui-ci veille à ce que les documents soient 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. Les numéros d'identification ont été attribués afin de préserver la confidentialité des élèves. Les noms des élèves dans le *Formulaire de suivi des élèves* ont été utilisés pour faciliter le déroulement de l'évaluation dans les écoles. Des formulaires d'ouverture de session supplémentaires et, au besoin, des tests sur papier ont été fournis afin de permettre la participation des nouveaux élèves inscrits après l'envoi de la *Liste des élèves* initiale.

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 que son intégrité ne soit compromise. Par exemple, des versions en gros caractères étaient offertes tant pour l'évaluation sur papier que pour celle en ligne en réponse aux besoins de ces élèves.

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 avaient accès à des ordinateurs et où ils 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 ou une page avec les coordonnées d'ouverture de session 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 a ramassé tous les documents d'évaluation et les a rangés en lieu sûr, afin que le matériel d'évaluation reste confidentiel. Les ordinateurs portables empruntés ont été remis dans leur emballage et retournés au CMEC ou envoyés à une autre école dans la même province.

De pair avec les coordonnatrices et coordonnateurs des écoles, les coordonnatrices et coordonnateurs des provinces ont également joué un rôle important 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. Ces coordonnatrices et coordonnateurs 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 ou que l'adaptation soit activée dès que l'élève lance l'évaluation. 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 (pour le test sur papier seulement), 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 pause 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

Les questionnaires pour la direction de l'école et le personnel enseignant ont été envoyés en ligne et étaient accessibles à l'aide d'un code unique assorti d'un mot de passe. Le *Questionnaire de l'école* était généralement rempli par la directrice ou le directeur de l'école. Les enseignantes et enseignants de mathématiques des classes sélectionnées devaient remplir le *Questionnaire du personnel enseignant*. 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 mathématiques associés aux élèves sélectionnés étaient priés de remplir le questionnaire (à raison d'un questionnaire par enseignante ou enseignant).

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. Toutes les réponses inscrites dans les questionnaires étaient confidentielles. L'utilisation des noms des enseignantes et enseignants avait uniquement pour but de faire le lien entre le numéro d'identification du *Questionnaire de l'élève* et celui du *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 mathématiques, en lecture 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 11 codes de participation :

- 1 = L'élève est absent.
- 2 = L'élève a participé pendant une séance prévue.
- 2A = L'élève a participé pendant une séance prévue avec une mesure d'adaptation.
- 3 = L'élève a participé pendant une séance de rattrapage.
- 4 = L'élève est exempté par l'école.
- 5 = L'élève est exempté parce que des mesures d'adaptation appropriées ne pouvaient pas être offertes.
- 6 = L'élève n'est plus inscrit dans cette école/classe.
- 7 = Les parents ou les élèves ne veulent plus se soumettre à l'évaluation.
- 8 = L'élève n'est pas en 8^e année/2^e secondaire.
- 9 = L'élève reçoit son enseignement à la maison.
- 10 = La feuille de réponse et le livret n'ont pas été remis; seulement les données du questionnaire.
- 11 = L'élève a répondu à moins de trois items par domaine et n'a pas rempli au minimum la première section (Section 1) du questionnaire contextuel.

Il y avait également trois codes d'exemption :

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 :

- temps supplémentaire : 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 mathématiques et de sciences seulement (dans les cas où il faut lire en entier le test de mathématiques ou de sciences, 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 du rapport public du PPCE 2019 (O'Grady, Houme et coll., 2021, p. 217).

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 1 (absence) et 2 + 2A (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 :

$$\frac{(2 + 2A)}{(1 + 2 + 2A)} \times 100$$

Si le taux de participation des élèves était inférieur à 85 p. 100, une séance de rattrapage devait être organisée. 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 1 (absence) à 3 (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 ont rempli le *Rapport de la coordonnatrice ou du coordonnateur de l'école*. Ils ont aussi rempli la *Liste des documents expédiés* et indiqué le nombre d'exemplaires de chaque type de 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 codage

La séance de codage des items de l'étude principale a eu lieu à Ottawa, pendant deux semaines en juillet 2019. Tous les items de mathématiques, de lecture et de sciences ont été codés par des enseignantes et enseignants de ces domaines. En tout, 102 codeuses et codeurs y ont participé, à la fois les francophones et les anglophones, plus deux chefs de codage par domaine.

Au total, environ 33 000 livrets d'évaluation ont été codés à l'aide du Système central de codage (SCC) en ligne, dont environ 8 000 en français et 25 000 en anglais.

Codage

Tous les items à réponse construite ont été codés par des membres du personnel éducatif, car ces items exigeaient un certain jugement de leur part et des connaissances de la matière. Les codeuses et codeurs se sont appuyés sur les descriptions des guides de codage pour attribuer divers codes aux réponses des élèves et les ont enregistrés dans le SCC. À la fin de la séance de codage, une base de données a été créée avec toutes les données de l'évaluation et des questionnaires.

Manuel de la codeuse ou du codeur

Avant la séance de codage, les codeuses et codeurs se sont vu fournir le *Manuel de la codeuse ou du codeur* contenant des informations sur la logistique de la séance de codage et décrivant les responsabilités du personnel du CMEC, des chefs de codage et des codeuses et codeurs. Ce guide fournissait également des informations sur le traitement des cas spéciaux, comme les points de vue tendancieux de la part de la codeuse ou du codeur ou les suspicions de triche. Le *Formulaire de commentaires de la codeuse ou du codeur*, qui devait être rempli à la fin de la séance de codage, 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 codage

Les chefs de codage se sont rencontrés quelques jours en juin pour préparer la séance de codage. 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 codeuses et codeurs. Lors de la préparation du matériel de formation, les chefs de codage ont choisi des échantillons de travaux d'élèves pouvant être utilisés comme exemples ou dans les documents de formation. 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 codage avaient pour responsabilité d'assurer la formation des chefs de table et de garantir le bon déroulement de la séance de codage.

Chefs de table

Les chefs de table dirigeaient chacun une table de six à huit codeuses et codeurs. Ils avaient été formés par les chefs de codage. Leur rôle était de former les codeuses et codeurs 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 codeuses et codeurs

L'ensemble des codeuses et codeurs, y compris les chefs de table, ont suivi une formation sur les guides de codage pour les mathématiques, la lecture ou les sciences, selon le rôle qui leur était attribué, avant de se lancer dans le codage des copies des élèves. Avant la séance de formation, les chefs de codage 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 trousseaux de formation ont ensuite été utilisés par les codeuses et codeurs pour s'exercer au codage et pour mieux intérioriser le système de codage. Au début, les codeuses et codeurs se sont mis par deux pour coder ensemble les réponses de élèves, jusqu'à ce que le codage soit conforme au guide. À la fin de la formation, lorsque les codeuses et codeurs étaient capables d'appliquer les codes de façon cohérente, ils sont passés à un travail individuel de codage, jusqu'à ce que toutes les réponses des élèves soient codées. Les codeuses et codeurs ont suivi la formation sur un seul et même item et ont codé toutes les réponses des élèves à cet item avant de suivre la formation sur l'item suivant. Tout au long du codage, les chefs de table ont procédé à une vérification aléatoire des codes donnés par chaque codeuse et chaque codeur, pour s'assurer de la conformité constante aux guides de codage. En cas de problème avec une question en particulier, la formation était reprise avec la codeuse ou le codeur ou avec l'ensemble du groupe et, dans quelques cas, l'item était codé de nouveau.

Les tables se sont vu attribuer soit des réponses d'élèves en anglais soit des réponses d'élèves en français. Les tables de codeuses et codeurs bilingues, qui pouvaient aider soit l'équipe anglophone soit l'équipe francophone dans le travail de codage des items, ont été affectées soit aux réponses d'élèves en anglais soit aux réponses d'élèves en français, selon le secteur qui avait plus de réponses ou qui avançait plus lentement dans le codage.

Fidélité du codage

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

Examens de fidélité

Dans le cadre d'une séance de codage, il est toujours important d'instaurer les procédures nécessaires pour s'assurer que les codeuses et codeurs 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 codage, le personnel du CMEC a constitué pour chaque item plusieurs ensembles de réponses choisies au hasard pour faire des examens de fidélité. L'ensemble d'items a ensuite été distribué aux chefs de codage pour qu'ils en fassent le codage. Leurs réponses ont été retournées au personnel du CMEC, qui les a saisies dans le SCC à des fins de comparaison par rapport aux réponses des codeuses et codeurs lors de la séance de codage. Si les chefs de codage 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 codeuses et codeurs. Les examens de fidélité ont été exécutés pour l'ensemble des codeuses et codeurs francophones ou anglophones dans les trois domaines d'évaluation. Le but de ces examens était de surveiller la cohérence du codage 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 codage, tout le monde s'arrêtait de coder et codaient les mêmes échantillons de travaux d'élèves à l'aide du SCC.
- Les codes attribués par les codeuses et codeurs étaient comparés aux données de référence (fournies par les chefs de codage).
- Les résultats étaient immédiatement communiqués au chef de codage.
- Les chefs de codage faisaient le bilan avec l'ensemble du groupe ou avec des codeuses ou codeurs à titre individuel.
- Si la cohérence était inférieure à 80 p. 100 pour une question spécifique, les individus ou groupes de codeuses et codeurs suivaient une nouvelle formation et les réponses des élèves étaient codées 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 codeuses et codeurs, c'est-à-dire qu'ils ont permis de vérifier si ces derniers avaient attribué les mêmes codes que les spécialistes pour les items. Pour chaque examen de fidélité, il y avait un pourcentage du degré d'accord calculé pour chaque codeuse et chaque codeur et pour chaque item. Il était attendu que le degré d'accord entre les spécialistes (chefs de codage) et les codeuses et codeurs soit d'environ 85 p. 100. Si le résultat global des examens de fidélité était faible pour telle ou telle question, alors le groupe suivait une nouvelle formation et les réponses déjà codées étaient vérifiées à nouveau par les chefs de codage, ou bien le codage de la question recommençait. Si le résultat des examens de fidélité était faible pour des codeuses ou codeurs ou pour des tables spécifiques, alors les

chefs de table offraient une nouvelle formation à l'individu ou au groupe concerné avant de passer au codage. Les items précédemment codés par ces codeuses ou codeurs étaient vérifiés.

Le pourcentage du degré d'accord par codeuse ou codeur 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 codage, tous les pourcentages obtenus pour chaque examen de fidélité de chaque codeuse et chaque codeur ont été compilés, ce qui a constitué le degré d'accord total en pourcentage. Les résultats montrent que la majorité des codeuses et codeurs ont obtenu un degré 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.

Fiabilité entre membres du groupe de codage (codage multiple)

Le codage multiple est un indicateur de contrôle de la qualité pour le codage des réponses des élèves en mathématiques, en lecture et en sciences. Pour chaque item, 100 réponses choisies au hasard ont été attribuées à chaque codeuse et chaque codeur. Les chefs de codage ont mesuré le degré d'accord des codeuses et codeurs et ont ainsi pu détecter les problèmes tout au long de la procédure de codage pour chaque item, puis ils ont formé de nouveau les codeuses et codeurs, individuellement ou en groupe, selon les besoins.

Fiabilité des tendances

La fiabilité des tendances est un indicateur de contrôle de la qualité servant à estimer le niveau d'accord entre les codeuses et codeurs en mathématiques, en lecture et en sciences pour les items d'ancrage à l'évaluation du PPCE 2019, du PPCE 2016 et du PPCE 2013. Quatre items pour la lecture, huit items pour les mathématiques et six items pour les sciences étaient communs aux trois éditions. Pour chaque item, 385 réponses d'élèves du PPCE 2016 ont été codées de nouveau en 2019. Ces réponses ont été tirées des livrets 1, 2 et 4. Le livret 3 n'était pas requis, en raison du recoupement des items des différents livrets. Les réponses des élèves ont été scannées et saisies dans le SCC. Ainsi, il était impossible pour les codeuses et codeurs de savoir si les réponses qu'ils codaient étaient celles de 2016 ou celles de 2019, celles de 2019 comportant également des réponses données par des élèves qui avaient rempli le test sur papier. Le codage pour la fiabilité des tendances s'est fait tout au long de la séance de codage principale.

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 ou soumises directement au Secrétariat du CMEC. Les codeuses et codeurs ont fait des commentaires lors de la séance de codage. 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. Par exemple, le guide encourageait 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.

Les coordonnatrices et coordonnateurs des provinces devaient également résumer les problèmes rencontrés par les écoles lors du déroulement de l'évaluation. De façon générale, très peu d'écoles ont éprouvé des problèmes techniques lors du déroulement du test.

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. Certaines écoles ont aussi fait part de préoccupations concernant la spécificité et la clarté des informations relatives au déroulement de 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, et 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 pour les tests sur papier.

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.

Équivalence et examen des items

Pour que l'évaluation sur papier soit le plus similaire possible à l'évaluation en ligne, les items ont été élaborés, traduits et examinés sur papier en premier, avant d'être importés sur la plateforme du test. Les items à réponse choisie ont été conçus pour que les élèves puissent choisir à l'aide de boutons radio une réponse à choix multiple ou à choix dichotomique (p. ex., vrai ou faux, oui ou non). Plusieurs items créés lors des évaluations précédentes du PPCE ainsi que de nouveaux items de ce cycle nécessitaient une forme ou un autre d'animation. Il s'agissait, par exemple, de faire glisser les réponses pour les placer en ordre, de créer un graphique linéaire ou de tracer des coordonnées. Pour les questions qui exigeaient la démonstration de la démarche de l'élève, un clavier mathématique était intégré dans le système. En outre, ce dernier comportait une calculatrice, un surligneur et deux onglets, sur lesquels les élèves pouvaient cliquer pour accéder à des références (telles que des formules mathématiques) et un glossaire. Les activités d'examen des items ont été menées à l'aide d'un système d'examen en ligne sécurisé, mis au point par le fournisseur de la plateforme

Configuration informatique requise et déploiement du test

Le matériel de base nécessaire pour le déploiement du test était un ordinateur par élève. Les ordinateurs étaient disposés de telle sorte que la supervision pouvait être effectuée par une seule administratrice ou un seul administrateur de test et que les élèves ne pouvaient pas facilement voir l'écran de leurs camarades. La configuration des appareils électroniques compatibles est présentée dans le tableau 7.1.

TABLEAU 7.1 Configuration informatique requise pour le PPCE en ligne

| Appareil électronique | Exigences |
|---------------------------|--|
| Microsoft | Windows 7, 8.1, ou 10 |
| Macintosh (Mac) | OSX 10.10–10.13 |
| Chromebook | ChromeOS 62+ Environnement de gestion requis |
| iPad | iOS 9.3.5+ Taille minimale de l'écran : diagonale de 7,9 pouces |
| Samsung ou autres Android | 6.0+ Taille minimale de l'écran : diagonale de 9,7 pouces |

Pour déterminer si un ordinateur convenait au test assisté par ordinateur, le navigateur du PPCE – AWIS Secure Access (accès sécurisé) – a été téléchargé sur tous les appareils électroniques utilisés dans chaque école pour le déploiement du PPCE. Il comprenait les paramètres par défaut appropriés pour chaque système d'exploitation. Lorsqu'il était activé, il empêchait les élèves d'accéder à Internet et à d'autres programmes pendant toute la durée de l'évaluation. Les élèves pouvaient utiliser leur propre appareil électronique, si les politiques du conseil, de la commission ou du district scolaire

le permettaient. Une fois le navigateur AWIS Secure Access installé, chaque appareil a été testé au moyen de l'outil d'exercice en ligne du PPCE, qui confirmait que l'évaluation fonctionnait bien. Les écoles qui ne disposaient pas d'un nombre suffisant d'appareils compatibles ont emprunté les ordinateurs portables du CMEC pour mener l'évaluation du PPCE. Celles qui n'avaient pas un accès adéquat à Internet ont reçu du CMEC des appareils WiFi portatifs nécessitant une couverture cellulaire. Les écoles pouvaient faire appel au service de soutien technique du PPCE pour la vérification des appareils ainsi que pendant tout le déroulement de l'évaluation.

Les tests se sont déroulés de façon sécuritaire sur Internet et étaient protégés grâce au chiffrement SSL de 128 à 256 bits de Verisign/Symantec. Tous les serveurs étaient protégés à l'aide d'un certificat Verisign/Symantec à protocole de sécurité inversé, similaire à celui utilisé pour les services bancaires en ligne. Les élèves avaient accès au test par l'entremise d'un lien sur le site Web du CMEC. Un formulaire d'ouverture de session a été remis pour chaque classe, avec un code d'accès et un mot de passe pour chaque élève. En cas de défaillance du système ou de panne, le système reprenait automatiquement le test sur l'appareil individuel, à l'endroit même où l'élève était au moment de l'interruption.

Les adaptations offertes aux élèves étaient similaires pour les tests en ligne et sur papier; avant l'évaluation, les écoles ont identifié les élèves qui avaient besoin d'un autre format. Les élèves qui ont passé le test en ligne pouvaient choisir des fichiers vocaux préenregistrés ou un logiciel de lecture de texte approuvé (accès non visuel à l'ordinateur portable [logiciel libre et gratuit NVDA], VoiceOver d'Apple ou ChromeVox de Google). Les logiciels de lecture de texte qui exigeaient l'accès à un programme externe n'étaient pas compatibles avec le navigateur AWIS Secure Access, car l'accès à Internet était restreint pendant l'évaluation.

Une fois le test terminé, les écoles ont rempli un formulaire de rapport de séance pour donner des renseignements sur les problèmes techniques, les interventions et les éventuelles adaptations utilisées pour les élèves.

Saisie des données et codage des réponses des élèves

Les réponses étaient sauvegardées sur les serveurs chaque fois que l'élève passait à la question suivante. Cette sauvegarde se produisait chaque fois que l'élève utilisait les boutons « Suivant » et « Précédent » et consultait les écrans « Révision des instructions ». Lorsque l'élève prenait plus de temps pour répondre, le système sauvegardait automatiquement les données toutes les minutes.

Après l'évaluation, les réponses construites de l'évaluation en ligne ont été transférées dans le Système central de codage (SCC) en ligne. Les réponses construites des livrets papier ont été scannées et saisies dans le système en ligne.

Pour les items à réponse choisie, les items à réponse numérique et les items à réponse glisser-déposer, le codage était automatique. Les réponses des élèves aux items à réponse ouverte étaient accessibles sur le SCC pour un codage humain. Pour chaque domaine, les codeuses et codeurs accédaient au système à l'aide d'un code et d'un mot de passe uniques. Ils pouvaient voir la liste des items à coder pour chaque scénario. Ils n'avaient qu'à cliquer sur le bouton radio du chiffre correspondant à la description du code fournie dans le *Guide de codage*. Dans l'espace réservé pour la réponse, les codeuses et codeurs pouvaient ajouter des commentaires et envoyer la réponse de l'élève au chef de codage, s'ils détectaient un élève à risque, un langage inapproprié ou un possible cas de triche.

Les items à coder ont été répartis à l'aide du logiciel du SCC. La qualité du codage a été vérifiée de différentes façons. Les réponses des élèves ont été distribuées au hasard aux codeuses et codeurs affectés à chaque ensemble d'items. Le codage multiple a été appliqué à un sous-ensemble de réponses, pour lequel les codeuses et codeurs avaient reçu le même ensemble de réponses des élèves, réparties de manière aléatoire parmi toutes les réponses qui leur avaient été confiées. Des rapports sur le codage multiple ont été produits pour chaque item et chaque codeuse et codeur. Les écarts entre les items soumis au codage multiple et les réponses jugées difficiles ou inappropriées par les codeuses et codeurs ont été communiqués au chef de table ou au chef de codage. Le SCC a fourni en temps réel des renseignements sur le processus de codage, par item et par codeuse et codeur.

Les tests de fiabilité constituent un autre mécanisme d'assurance de la qualité; environ deux tests de fiabilité ont été menés pour chaque item. Les chefs de codage ont choisi des exemples de réponses à soumettre aux tests de fiabilité, qui ont été distribués à l'aide du SCC. Les rapports de fiabilité produits portaient sur chaque item et chaque codeuse et codeur. Ces tests de fiabilité ont aidé à déterminer si les codes avaient été attribués correctement. Cette mesure a permis de cerner tôt les problèmes de codage et a été suivie d'une discussion et, au besoin, d'une reprise de la formation.

Pendant la séance de codage, les chefs de table ont vérifié de façon ponctuelle le travail quotidien des codeuses et codeurs. Pour ce faire, ils devaient notamment examiner les codes attribués aux réponses. La règle générale était qu'au moins 10 réponses par item devaient être vérifiées de façon ponctuelle. Si une codeuse ou un codeur éprouvait des doutes quant au code à attribuer à une réponse, il pouvait l'indiquer pour qu'elle soit automatiquement envoyée au chef de table ou au chef de codage afin qu'il se prononce.

Une fois la collecte des données terminée, celles-ci ont été transférées au CMEC à l'aide d'un protocole de transfert de fichiers sécuritaire.



Le PPCE 2019 marque le début de la transition de l'évaluation sur papier à une évaluation en ligne. Le mode d'administration a fait l'objet d'une étude pilote, et une étude sur le mode d'administration a été réalisée à la fois lors de la mise à l'essai et de l'étude principale du PPCE 2019 pour déterminer si les items d'ancrage, soit ceux utilisés dans les évaluations précédentes du PPCE, fonctionnaient de la même façon dans leur version en ligne que dans leur version papier.

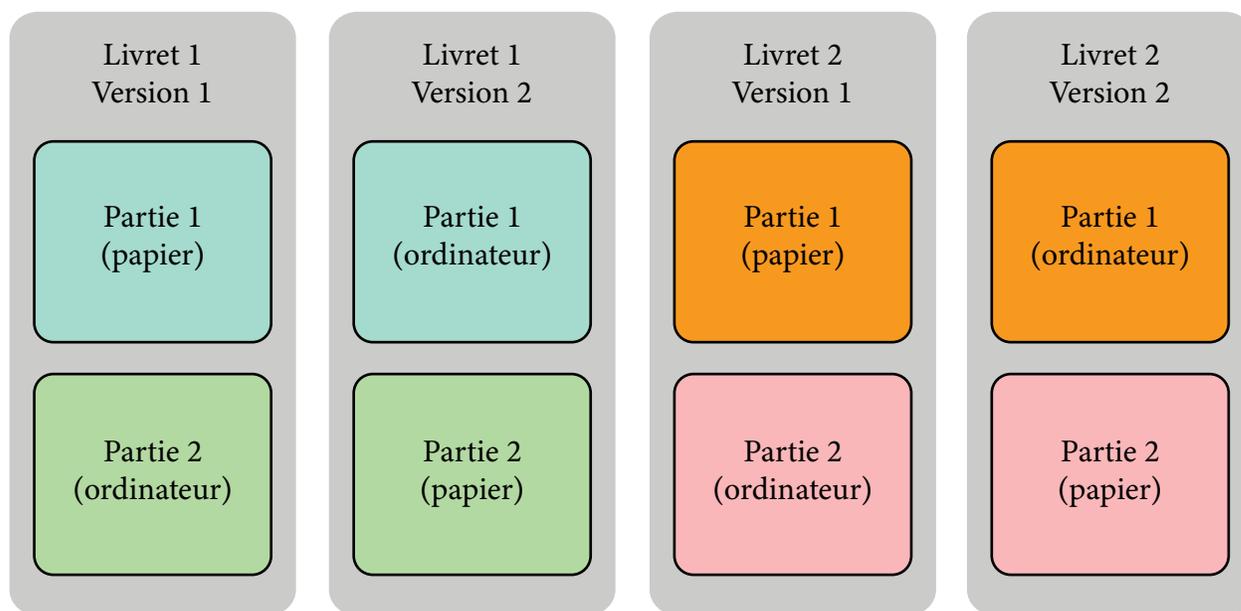
L'étude pilote a été réalisée dans 43 écoles en novembre 2016. Deux fournisseurs ont été choisis pour cette étude. Le premier a utilisé un test par navigateur Web, le second, un test par application privée (kiosque). Des ordinateurs PC et Mac ainsi que des ordinateurs portables Chromebook et des tablettes Android, entre autres appareils, ont été utilisés pour l'étude. Celle-ci a démontré que le recours à une plateforme en ligne entraînait moins de défis technologiques dans les écoles. L'étude pilote en ligne a déterminé que toutes les écoles seraient en mesure de répondre aux exigences technologiques minimales nécessaires au déroulement de l'évaluation du PPCE en ligne.

Aperçu de l'étude sur le mode d'administration dans le cadre de la mise à l'essai

La première étude sur le mode d'administration a été effectuée au printemps 2017 auprès d'un échantillon distinct de 156 écoles au moyen d'une plateforme de test en ligne. Un sondage sur l'état de préparation sur le plan technologique a été envoyé à ces écoles avant le déploiement du test, et un exemple de test leur a été fourni pour veiller à ce que l'équipement soit configuré de façon adéquate.

Les livrets de l'étude sur le mode d'administration comprenaient des items qui ont été utilisés comme items d'ancrage dans les évaluations antérieures du PPCE dans les trois domaines (mathématiques, lecture et sciences). Deux livrets parallèles ont été conçus, lesquels comportaient 33 ou 34 items divisés en deux sections. L'étude sur le mode d'administration était un test de 90 minutes effectué en deux parties, les élèves répondant d'abord à une section sur papier, puis à une section en ligne, ou vice-versa, comme le montre la figure 8.1.

FIGURE 8.1 Conception des livrets pour l'étude sur le mode d'administration de la mise à l'essai



Les résultats des élèves ont été agrégés à l'échelle du test afin de comparer les écarts de rendement entre les élèves qui ont répondu aux mêmes items du test, mais avec un mode d'administration différent. Par exemple, le rendement des élèves pour le livret 1, partie 1, version 1 (papier) a été comparé à celui des élèves qui ont répondu à la version 2 (ordinateur).

L'étude sur le mode d'administration était conçue selon un modèle à un groupe avec compensation. Le fait d'alterner le mode d'administration a permis d'assurer que l'ordre dans lequel l'évaluation sur papier et l'évaluation en ligne étaient soumises aux élèves n'affecterait pas les scores des élèves. À cet égard, la conception facilite une comparaison directe entre les deux modes d'administration. Par exemple, dans un livret donné, la partie 1 (papier) est directement comparable à la partie 1 (ordinateur), et la partie 2 (papier) est comparable à la partie 2 (ordinateur). Dans chaque cas, les groupes sont constitués de façon aléatoire et satisfont à cette exigence.

L'analyse du fonctionnement différentiel des items (FDI) a révélé un écart significatif entre les deux modes d'administration, particulièrement en lecture. En lecture, des scores moyens supérieurs ont été observés pour les items soumis aux élèves en ligne dans le livret 1, mais l'inverse a été observé pour le livret 2. En mathématiques, les résultats moyens étaient supérieurs dans les deux livrets soumis aux élèves sur papier. Des scores similaires ont été observés entre les deux modes d'administration en sciences.

Pour mieux comprendre ces écarts, l'analyse du fonctionnement différentiel du test (FDT) a été effectuée afin de déterminer si un biais se produit à l'échelle du test pour chaque domaine. Les données pour tous les livrets ont été combinées, et trois analyses du FDT ont été effectuées (une par domaine). Un logiciel de système d'analyse du fonctionnement différentiel des items (FDI) a été utilisé pour réaliser l'analyse du FDT. Ce système a examiné l'écart de l'effet du FDI comme mesure du FDT, et le FDI a été estimé au moyen du chi carré de Mantel-Haenszel et du chi carré de Mantel. Les résultats de l'analyse du FDT montrent d'importants écarts de l'effet du FDI (important biais du FDT) pour les mathématiques et la lecture, et un faible écart de l'effet du FDI (faible biais du FDT) pour les sciences.

Les spécialistes en matière d'évaluation et de programmes d'études provinciaux ont examiné les résultats et ont recensé les items du domaine de la lecture pour lesquels la présentation de la mise en situation et des items était différente entre les deux modes d'administration. Les mises en situation en lecture sont présentées avant les items dans les tests sur papier, de façon que l'élève lise normalement le passage avant de voir les items. Dans le test en ligne, les items ont été présentés à gauche sur l'écran, et le passage apparaît à droite, donc les élèves ont peut-être pensé qu'ils devaient trouver l'information précise dans le passage. De plus, les données ont été présentées selon la technique d'échantillonnage matriciel de façon que tous les items d'un domaine soient traités comme un « test unique ». Cependant, chaque « test » comprend seulement de petits lots d'items pour chaque élève.

Une deuxième étude sur le mode d'administration était prévue pour l'étude principale afin d'analyser plus en profondeur l'écart de fonctionnement des items entre les deux modes d'administration au moyen d'un plus grand échantillon et de régler les problèmes de présentation des items, au besoin.

Aperçu de l'étude sur le mode d'administration dans le cadre de l'étude principale

Au cours de l'étude principale du PPCE 2019, 29 926 élèves ont participé à l'évaluation – 26 933 élèves ont fait l'évaluation du PPCE 2019 en ligne, alors qu'un échantillon distinct composé de 2 993 élèves ont fait l'évaluation sur papier. Les versions en ligne et sur papier des quatre livrets étaient aussi similaires que possible. Des différences ont été rapportées en ce qui a trait à la présentation et aux outils. Par exemple, dans la version en ligne, les élèves ont utilisé des boutons radio pour répondre aux questions à réponse choisie, ont pu consulter des liens vers des termes de mathématiques et de sciences en français et en anglais et ont eu accès à une calculatrice et à des outils de création de graphiques pour des questions précises. Sur la version papier, des pages détachables ont été utilisées pour les feuilles de réponses des questions à réponse choisie et pour les termes de mathématiques et de sciences traduits. Le *Questionnaire de l'élève* a été soumis après le test cognitif dans les deux versions du test.

Le programme IRTPRO a été utilisé pour les analyses du FDI. Ce programme a recours au test de Wald par Lord, avec des matrices de variance-covariance d'erreurs de paramètres d'items précises, calculées au moyen de l'algorithme SEM (Supplemented Expectation Maximization [algorithme espérance/maximisation amélioré]) (Cai, 2008; Vector Psychometric Group, 2020). Les items ont été regroupés par matière (mathématiques, lecture et sciences). Des analyses du FDI ont été réalisées pour tous les items dans chaque matière. Les paramètres de difficulté ont d'abord été comparés, en corrélant les paramètres b entre les modes. Ensuite, les résultats du FDI obtenus du test de Wald ont été examinés, et pour les items affichant un FDI important, le mode ayant présenté un « avantage » a été noté. Les scores des élèves basés sur les calibrations de la théorie des réponses aux items (TRI) ont aussi été obtenus pour comparer les écarts de rendement; pour faciliter l'interprétation, les scores ont été rapportés à une échelle avec une moyenne de 500 et un écart-type de 100.

Les analyses préliminaires montrent qu'environ la moitié des items présentent un FDI; toutefois, il y a à peu près autant d'items qui favorisent la version en ligne que d'items qui favorisent la version papier.

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 2019, les mathématiques ont été le domaine principal du PPCE pour la deuxième fois. Le cadre d'évaluation a été mis à jour et le PPCE est devenu un test en ligne, 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 2020 à Toronto. Les réunions ont été divisées en trois séances : une séance de formation des chefs d'une journée; une séance d'établissement des normes de deux jours; et une séance de rédaction d'une journée, pour réviser les descripteurs des niveaux de rendement.

L'établissement des normes avait pour but de définir les niveaux de rendement pour l'évaluation en mathématiques 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.

La séance d'établissement des normes a été animée par M. Lennie Comeau, ministère de l'Éducation de la Nouvelle-Écosse, appuyé par deux collègues de son ministère et le personnel du Secrétariat du CMEC. Les participants ont fait les tests, les ont codés, 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, un pour chaque secteur linguistique, ayant une certaine expertise en mesure et en évaluation, et en ce qui concerne le contenu du programme de mathématiques pour la 8^e année/2^e secondaire. Le comité d'établissement des normes se composait de 27 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 27 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 2010, année pour laquelle les mathématiques ont été le domaine principal pour la première fois, ont été examinés. Il s'agissait d'énoncés décrivant ce que les élèves des quatre 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.

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 2019 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 mathématiques auxquels les élèves pouvaient répondre par une réponse brève qui était codée 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, à laquelle un crédit partiel pouvait être donné.

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 ou ses livrets d'origine, sa classification par sous-domaine mathématique, 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 l'expression « 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, entre les niveaux 2 et 3 et entre les niveaux 3 et 4.

Procédure d'établissement des normes

Au total, 27 participantes et participants de l'ensemble des provinces ont pris part à la séance d'établissement des normes. Les participantes et participants ont été affectés à l'une des cinq tables prévues, soit deux tables francophones ou trois tables anglophones. Chaque table avait un chef et cinq ou six 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 codage des items à choix multiples et les guides de codage 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 quatre 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 coder leurs réponses, de façon à se faire une meilleure idée des descripteurs des niveaux de rendement.

La méthode du signet nécessite trois rondes, dont les deux premières sont des rondes de pratique, pour que les participantes et participants se familiarisent avec la procédure. La première ronde de placement des signets a eu lieu avant la fin de la première journée : les participantes et participants ont d'abord examiné indépendamment les uns des autres chaque item dans le LIO. Chaque participante ou participant choisissait un point de coupure – c'est-à-dire qu'il plaçait 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. Puis les groupes ont discuté 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, toutes les réponses de l'ensemble des participantes et participants ont été enregistrées dans un fichier Excel, avec les numéros de page du LIO indiquant les trois 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, entre le niveau 2 et le niveau 3 et entre le niveau 3 et le niveau 4.

La deuxième journée a commencé par une discussion en séance plénière sur la première ronde de placement des signets. Les résultats ont été affichés avec un graphique illustrant tous les numéros de page du LIO notés par les spécialistes aux trois 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. 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.

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 trois points de coupure et fournissant la justification de leur choix en groupe. L'objectif de la deuxième ronde était d'obtenir un ensemble de résultats qui soit plus cohérent que celui constitué lors de la première ronde. Les résultats ont été compilés et communiqués 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 points de coupure que lors de la première ronde. Pour la deuxième ronde, 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 à chacun des niveaux de rendement. 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 9.1.

TABLEAU 9.1 Répartition des élèves par niveau de rendement en mathématiques

| Niveau de rendement | Niveau 1 | Niveau 2 | Niveau 3 | Niveau 4 |
|------------------------------|--------------|-----------|-----------|-------------|
| Score moyen en mathématiques | 385 ou moins | 386 à 497 | 498 à 644 | 645 ou plus |
| Pourcentage d'élèves | 11 | 39 | 42 | 8 |

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 scénarios et les questions étaient appropriés et équitables pour des élèves de 8^e année/2^e secondaire.

⁹ Comme il est décrit plus haut, la statistique thêta a été ajustée pour la probabilité de réponse de deux sur trois.

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 quatre niveaux de rendement. Ils ont actualisé, à partir de ces items, la description des connaissances et des habiletés caractérisant le rendement des élèves à chacun des quatre niveaux¹⁰. Selon les attentes pancanadiennes des programmes d'études en mathématiques, le niveau de rendement attendu des élèves de 8^e année/2^e secondaire est le niveau 2. Les élèves qui se situent au 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 4) et cette progression relative à la difficulté et à la complexité des tâches s'applique à la fois aux mathématiques globalement et à chaque sous-domaine de l'évaluation.

¹⁰ Ces descriptions figurent dans le rapport public sur le PPCE 2019 (O'Grady, Houme et coll., 2021).

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. L’objectif était de demander leur avis et leurs éventuels conseils pour appuyer les analyses.

Nettoyage des données

Lorsque les données ont été reçues après la séance de codage, 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 codés à l’aide du code de participation approprié. Cette correction s’appliquait également aux élèves qui n’avaient pas répondu à au moins trois items par domaine et qui n’avaient pas rempli au minimum la partie du questionnaire contextuel qui portait sur les antécédents (section 1).
- 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. Ces analyses ont 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, la clé a été corrigée, au besoin, et les données ambiguës ont fait l'objet d'un codage supplémentaire. Par exemple, si l'ambiguïté était liée à des erreurs d'impression, d'agencement 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.

Ce chapitre décrit l'analyse des données sur le rendement des élèves du PPCE 2019. Il fournit et décrit 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 2019 en mathématiques, en lecture et en sciences, et les résultats du PPCE 2007, 2010, 2013 et 2016
5. codage et mise à l'échelle des données sur le rendement du PPCE 2019
6. estimations des erreurs-types
7. présentation des résultats sur le rendement des élèves du PPCE 2019

Analyse préliminaire

L'analyse préliminaire s'inscrivait dans le 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. À ce stade du nettoyage, on a cerné les cas où les élèves ayant participé à l'évaluation en étaient exemptés (codes de participation attribués 4, 5 ou 6) : il pouvait s'agir d'une exemption

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

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 papier comportant un numéro d'identification qui ne correspondait pas à celui du *Formulaire de suivi des élèves* ou au fait que l'élève avait reçu un second code d'accès en raison d'un problème technique avec l'évaluation en ligne. 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.

Recodage des items

Avant de procéder aux analyses des données, les ensembles de données brutes du PPCE 2019 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 2019 comprenaient à la fois des réponses valables et des réponses non valables pour les items du test. Pour un item à choix multiples (CM) de l'évaluation sur papier, 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. Cette situation ne pouvait pas se produire dans l'évaluation en ligne. 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 ces items n'avaient pas été recodés, 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, notamment pour mesurer le comportement des leurres.

Données manquantes

Comme pour les autres évaluations à grande échelle, quatre types de données manquaient dans l'évaluation du PPCE 2019¹² :

- données manquantes en raison de l'échantillonnage de l'item (item non utilisé dans un livret);
- réponse manquante parce que l'élève n'a pas vu la question (sans objet);
- 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 un livret

Les items *non utilisés* dans un livret 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 des répondantes et répondants 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 traiter 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.

Items sans objet

Le code « sans objet » a été utilisé si la question avait été mal imprimée ou mal affichée sur un appareil numérique, de sorte qu'il était impossible pour l'élève d'y répondre. Par exemple, il pouvait y avoir eu une erreur d'impression ou de résolution d'écran rendant la question illisible. Le code « sans objet » n'a été utilisé que dans un petit nombre de cas, et ces derniers 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

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.

qui le précède immédiatement ou si l'élève n'a 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 2019 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 qu'avec 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.

Réponse non valable

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

Analyse des items

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

Analyse des items dans le cadre de la théorie classique des tests

L'objectif de l'analyse classique était de produire des statistiques pour un deuxième examen des items du PPCE 2019. Pour le domaine principal, c'est-à-dire les mathématiques, le premier examen s'est appuyé sur les données de la mise à l'essai. La mise à l'essai de l'évaluation en mathématiques comportait uniquement des nouveaux items. Dans l'étude principale, l'évaluation en mathématiques comportait à la fois des nouveaux items et des items d'ancrage, alors que pour les domaines secondaires, soit la lecture et les sciences, seuls des items d'ancrage ont été utilisés. Les items d'ancrage, tirés des évaluations précédentes, lorsque la matière était le domaine principal (lecture en 2007 et 2016, mathématiques en 2010 et sciences en 2013) ont été utilisés afin d'évaluer les

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. 199).

changements dans le rendement des items au fil du temps (ou d'une cohorte à l'autre) en fonction de l'estimation de leur niveau de difficulté. Les statistiques ont été examinées en vue de la sélection des items à inclure dans le PPCE 2019.

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

- difficulté des items;
- discrimination des items;
- statistiques spécifiques pour les items à réponse choisie (RC) (items à choix multiple, items vrai ou faux, items oui ou non, etc.);
- statistiques spécifiques pour les items à réponse construite (RCS);
- pourcentage d'élèves choisissant chaque option de réponse pour chaque item;
- pourcentage d'élèves n'atteignant pas l'item;
- pourcentage d'élèves omettant l'item;
- 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 codage 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 –, à la fois pour la version en ligne et la version papier du test. Pour ces domaines secondaires, le placement des items dans l'évaluation du PPCE 2019 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, et obtenu, était une corrélation modérément positive entre les items avec de bonnes propriétés de mesure et l'échelle, et ce, peu importe le mode d'administration. 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 bisériale 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 bisériale 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 bisériale 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 bisériale 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 2019

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 bisériale 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 bisériale de point positive pour plus d'un leurre dans les items à CM ou corrélations bisériales 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;

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.

- moins de 10 p. 100 des élèves obtenant la valeur du score pour un item à RCS;
- accord entre membres du personnel de codage de moins de 70 p. 100 pour la valeur du score pour un item à RCS.

Analyse de la TRI

L'analyse de la TRI a fait intervenir les aspects suivants : (1) évaluation de la dimensionnalité; (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. Le processus s'est achevé par une évaluation du fonctionnement différentiel des items (FDI) dans le cadre du traitement des données sur la validité.

Évaluation de la dimensionnalité du PPCE 2019

La dimensionnalité 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 lorsque des variables dichotomiques (ou polytomiques) sont traitées (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 θ 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 logistique à deux paramètres (2-PLM). Elle a ensuite été comparée au modèle logistique 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 RCS, 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 $S\text{-}\chi^2$. (IRTPRO ne produit pas et n'approuve pas la production de la courbe de réponse aux items empirique.) Les statistiques $S\text{-}\chi^2$ se fondent sur la différence entre les fréquences observées et les fréquences attendues dans les catégories de réponses selon les scores additionnés.

Invariance du PPCE 2019

Dans le cadre de l'étude sur le mode d'administration, le fonctionnement différentiel des items (FDI) pour le mode d'administration du PPCE 2019 sur papier et en ligne a été étudié. Plus particulièrement, l'analyse visait à déterminer dans quelle mesure les résultats découlant de la nouvelle formule, soit le mode d'administration par ordinateur, étaient comparables à ceux obtenus au moyen du mode d'administration sur papier, et donc à ceux des cycles précédents. À cette fin, il était important de déterminer si les items du PPCE 2019 présentaient des propriétés psychométriques similaires entre les modes d'administration. Les items d'ancrage ont présenté un intérêt particulier, puisqu'il était essentiel que ceux-ci fonctionnent de façon équivalente entre les modes d'administration afin de maintenir l'évaluation de la variation au fil du temps.

Cette recherche a été menée au moyen d'un ensemble d'analyses du FDI : deux processus d'évaluation du FDI fondés sur la TRI et une technique non fondée sur la TRI. La première analyse a eu recours à l'approche fondée sur le test de Wald intégrée dans IRTPRO (test de Wald d'IRTPRO) pour l'évaluation du FDI. Cai (2010) a indiqué que ces tests sont conçus selon le modèle de la proposition de Lord (1977), soit « avec des matrices de variance-covariance d'erreurs de paramètres d'items précises, calculées au moyen d'un algorithme espérance-maximisation complété » [*traduction libre*] (Vector Psychometric Group, 2020). Étant donné la grande envergure de l'évaluation du PPCE, il était attendu que cette approche donne lieu à un nombre considérable d'items présentant un FDI statistiquement significatif. Toutefois, le résultat de l'analyse de la première ronde a été utilisé pour poursuivre les recherches. Par conséquent, deux approches additionnelles ont été utilisées pour les évaluations subséquentes : l'analyse de régression logistique du FDI avec calcul de l'ampleur de l'effet (LRDIF_ES) et une détection du FDI par la régression ordinaire logistique (*lordif*) fondée sur le cadre de la TRI (Choi, Gibbons et Crane, 2016a). La LRDIF_ES a été effectuée au moyen d'une macro SAS pouvant traiter simultanément des items dichotomiques et polytomiques (Fu et Monfils, 2016). La détection du *lordif* a été réalisée à l'aide d'un progiciel statistique R du même nom pouvant aussi simultanément gérer des items dichotomiques et des items polytomiques (Choi, Gibbons et Crane, 2016b). Les deux approches ont l'avantage de calculer l'ampleur de l'effet, une statistique plus utile pour évaluer l'ampleur de l'écart relatif au fonctionnement de l'item. Elles permettent aussi d'examiner le FDI uniforme et non uniforme au moyen d'un seul test composite. Cette opération consiste à adapter et à comparer, pour chaque item, trois modèles présentés dans un ordre hiérarchique et ascendant. Ces modèles sont, du plus parcimonieux au plus complexe : le

modèle sans FDI, le modèle avec FDI uniforme et le modèle avec FDI non uniforme. Ces modèles seront ci-après nommés modèle 1, modèle 2 et modèle 3, respectivement. Le modèle 1 pose comme postulat qu'aucun FDI n'est présent; seul le trait latent prédit la probabilité du score d'un élève pour un item donné. Le modèle 2 teste une condition de l'effet principal; il pose comme postulat que, en plus du trait latent, la composition du groupe prédit indépendamment la probabilité du score d'un élève pour un item donné – ce qui signifie qu'un FDI est présent et est constant sur le continuum des niveaux de trait. Le modèle 3 ajoute une interaction entre les deux variables explicatives, soit que l'effet de la composition du groupe varie, en fonction du niveau de trait.

Outre ces attributs et fonctionnalités communes, la LRDIF_ES permet aux chercheuses et chercheurs d'inclure des variables explicatives externes au-delà de l'habileté et de la composition du groupe dans son modèle (Fu et Monfils, 2016). Pour le PPCE 2019, les variables contextuelles comme le niveau de scolarité des parents et le nombre de livres à la maison, lesquelles reflètent une mesure du statut socioéconomique, ont été incluses dans le modèle logistique à titre de variables externes. La LRDIF_ES comprend aussi « une option de production de graphique en forme de ligne pour chaque item et variable de regroupement, dans lequel le score moyen de l'item ainsi que les limites inférieures et supérieures de l'intervalle de confiance de 95 p. 100 du score moyen pour chaque groupe sont représentés en fonction de la variable dépendante (correspondance) » [*traduction libre*] (Fu et Monfils, 2016, p. 1). Alors que la LRDIF_ES repose sur le score du trait additionné, le *lordif* a aussi l'avantage d'intégrer la régression logistique et les techniques de la TRI dans un seul cadre. Ainsi, les paramètres des items et le trait latent sont estimés au moyen de techniques de la TRI. Pour respecter le modèle d'IRTPRO, le modèle de crédit partiel généralisé a été adapté pour les items polytomiques du PPCE 2019 plutôt que le modèle de réponse graduée par défaut déployé dans le *lordif*. Le *lordif* a aussi recours à la méthode d'équivalence de Stocking-Lord pour permettre l'étude de l'incidence du FDI sur la même mesure. De plus, il utilise une simulation de Monte Carlo pour dériver les valeurs limites empiriques pour diverses statistiques du FDI et mesures de la magnitude.

Il est possible de constater que ces approches constituent un bon complément pour une façon plus exhaustive d'évaluer la signification pratique des items avec FDI du PPCE 2019. Dans les deux processus, un test du rapport des vraisemblances a été utilisé à des fins de comparaison des modèles. Un écart significatif observé entre le modèle 1 et le modèle 2 indiquait la présence d'un FDI uniforme. Si le test du rapport des vraisemblances donnait lieu à un écart significatif entre le modèle 2 et le modèle 3, cela signifiait la présence d'un FDI non uniforme. Un FDI global ou total était défendable si le résultat du test du rapport des vraisemblances comparant le premier et le dernier modèle était significatif. Même si le test du rapport des vraisemblances est utile et communément utilisé à des fins de comparaison de modèles, il essuie les mêmes limites que la plupart des tests statistiques, puisque la taille de l'échantillon peut avoir une incidence. Pour les échantillons de grande taille, il a tendance à produire des résultats statistiquement significatifs pouvant être pratiquement négligeables. Par conséquent, les statistiques sur l'ampleur de l'effet ont été utilisées pour évaluer la magnitude des items avec FDI du PPCE 2019. À cette fin, l'évaluation a reposé sur le coefficient R^2 de Nagelkerke, le coefficient R^2 de McFadden et le coefficient de régression (β). Plus spécifiquement, l'ampleur de la variation, dans ces indices (ΔR^2 et $\Delta \beta$), d'un modèle plus parcimonieux à un modèle plus complexe, révélait la magnitude du FDI. Pour les items dichotomiques et polytomiques, la magnitude, en ce qui concerne R^2 , a été classifiée au moyen des critères recommandés par Educational Testing Service – ETS (service d'évaluation en éducation) dans le processus de Mantel-Haenszel (Fu et Monfils, 2016; voir aussi Choi, Gibbons et Crane, 2011; Jodoin et Gierl, 2001). Selon ETS, un item avec FDI tombe dans l'une des trois catégories présentées dans le tableau 11.1.

TABLEAU 11.1 Classification du FDI selon Educational Testing Service (ETS)

| Catégorie | Critères | Magnitude |
|-----------|--------------------------------|-----------------|
| A | $\Delta R^2 < 0,035$ | FDI négligeable |
| B | $0,35 \leq \Delta R^2 < 0,070$ | FDI modéré |
| C | $\Delta R^2 \geq 0,070$ | FDI important |

Pour $\Delta\beta$, soit la variation proportionnelle des estimés ponctuels pour le bêta (la variation proportionnelle de bêta), différents critères ont été proposés pour conclure à une incidence pratiquement significative d'un item avec FDI. Par exemple, Maldonado et Greenland (1993) ont proposé un seuil de 10 p. 100 de variation, alors que d'autres ont suggéré des seuils plus bas (p. ex., 5 p. 100) pour cette mesure de l'ampleur de l'effet (Crane, van Belle et Larson, 2004).

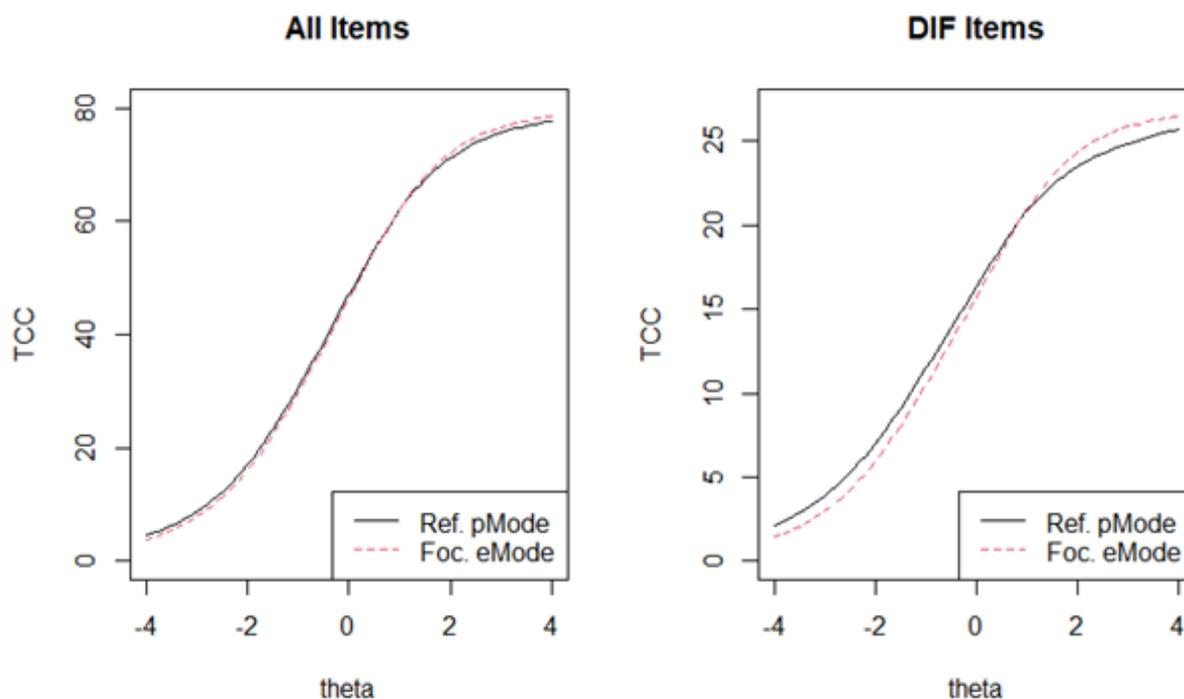
Les résultats du tableau 11.2 indiquent une magnitude allant de négligeable à modérée du fonctionnement différentiel des items du PPCE 2019 avec un FDI de modes. Les mesures de l'ampleur de l'effet ΔR^2 sont toutes situées sous 0,05. En fait, l'inspection des résultats par item (dont les résultats ne sont pas rapportés ici) a indiqué que seul un item a montré un FDI modéré ($\Delta R^2 = 0,036$). De façon similaire, toutes les valeurs pour les variations proportionnelles de β se trouvaient sous 0,01, sauf pour un item pour lequel la mesure totalisait 0,07. Il importe de signaler que les statistiques sur l'ampleur de l'effet s'alignent sur la représentation graphique de la magnitude du FDI, comme le montre la courbe caractéristique du test (CCT) illustrant le FDI. Pour les mathématiques, il y avait presque un chevauchement des CCT pour les modes d'administration en ligne et sur papier (voir la figure 11.1). Pour la lecture et les sciences, les CCT étaient très similaires et, donc, ne posaient pas de préoccupations. Il importe de noter que tous les items d'ancrage, dans les trois domaines, ont montré des amplitudes de l'effet marginales. Les conclusions ont été communiquées au comité consultatif technique. À la lumière de ces conclusions, il a été déterminé que les FDI détectés étaient pratiquement non significatifs. En d'autres termes, l'utilisation de données tirées du mode d'administration en ligne n'aurait pas de conséquences pour l'analyse des tendances et la comparaison avec les évaluations passées du PPCE.

TABLEAU 11.2 Fourchettes des mesures de l'ampleur de l'effet pour les items avec FDI selon le mode

| Domaine | ΔR^2 | $\Delta\beta$ |
|----------|--------------|---------------|
| MathN* | 0,001–0,036 | 0,001–0,070 |
| MathT* | 0,001–0,004 | 0,002–0,003 |
| Lecture | 0,001–0,004 | 0,001–0,003 |
| Sciences | 0,001–0,009 | 0,001–0,013 |

* MathN = nouveaux items de mathématiques; MathT = items d'ancrage de mathématiques

FIGURE 11.1 Détection du FDI selon le mode : courbes caractéristiques du test pour les items de mathématiques



L'analyse qui suit traite exclusivement des données du déploiement en ligne, puisqu'elles sont celles qui feront dorénavant l'objet des rapports. Avant de générer le trait d'habileté final sur lequel se pencher, une autre analyse du FDI a été effectuée pour la langue et le sexe. Celle-ci comprenait une évaluation de la mesure dans laquelle certains des items du PPCE 2019 affichaient des propriétés statistiques différentes (p. ex., le degré de difficulté) pour la langue et le sexe. Comme pour l'évaluation de l'invariance entre les modes d'administration, la détection du FDI du test de Wald, le processus LRDIF_ES et le processus lordif, tels que décrits ci-dessus, ont été appliqués aux données. Même si les résultats ont montré que certains des items présentaient un FDI selon la langue et le sexe, les ampleurs de l'effet n'étaient pas assez importantes pour supposer la non-invariance de la mesure. Les fourchettes des valeurs pour les mesures de l'ampleur de l'effet sont présentées au tableau 11.3 pour la langue et le sexe. Ces mesures représentent une magnitude allant de faible à modérée.

TABLEAU 11.3 Fourchettes des mesures de l'ampleur de l'effet pour les items avec FDI selon la langue et le sexe

| Groupes | Domaine | ΔR^2 | $\Delta\beta$ |
|---------|---------------|---------------|---------------|
| Langue | Mathématiques | < 0,001–0,044 | 0,001–0,065 |
| | Lecture | < 0,001–0,031 | < 0,001–0,045 |
| | Sciences | < 0,001–0,018 | < 0,001–0,015 |
| Sexe | Mathématiques | < 0,001–0,039 | 0,001–0,018 |
| | Lecture | < 0,001–0,027 | < 0,001–0,028 |
| | Sciences | < 0,001–0,018 | 0,001–0,010 |

Les CCT se chevauchaient presque parfaitement pour le genre, ou de très près pour la langue. Elles sont représentées dans les figures 11.2 et 11.3, respectivement.

FIGURE 11.2 Détection du FDI selon le sexe : courbes caractéristiques du test pour les items de mathématiques

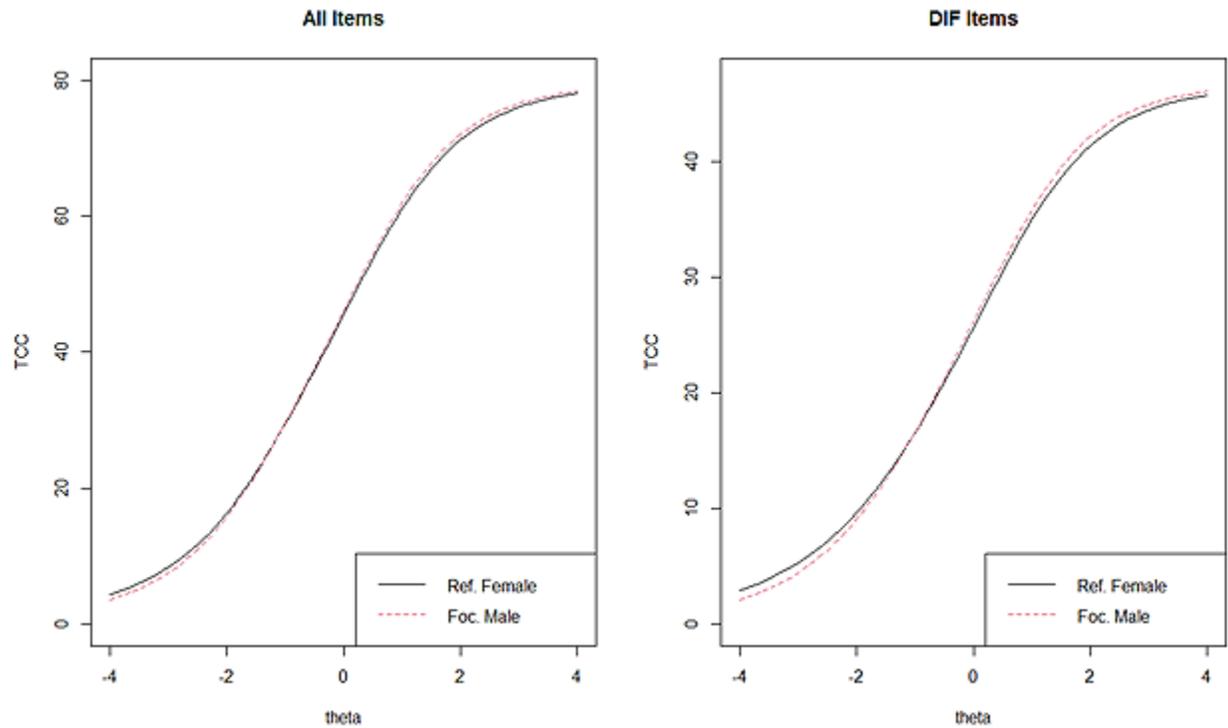
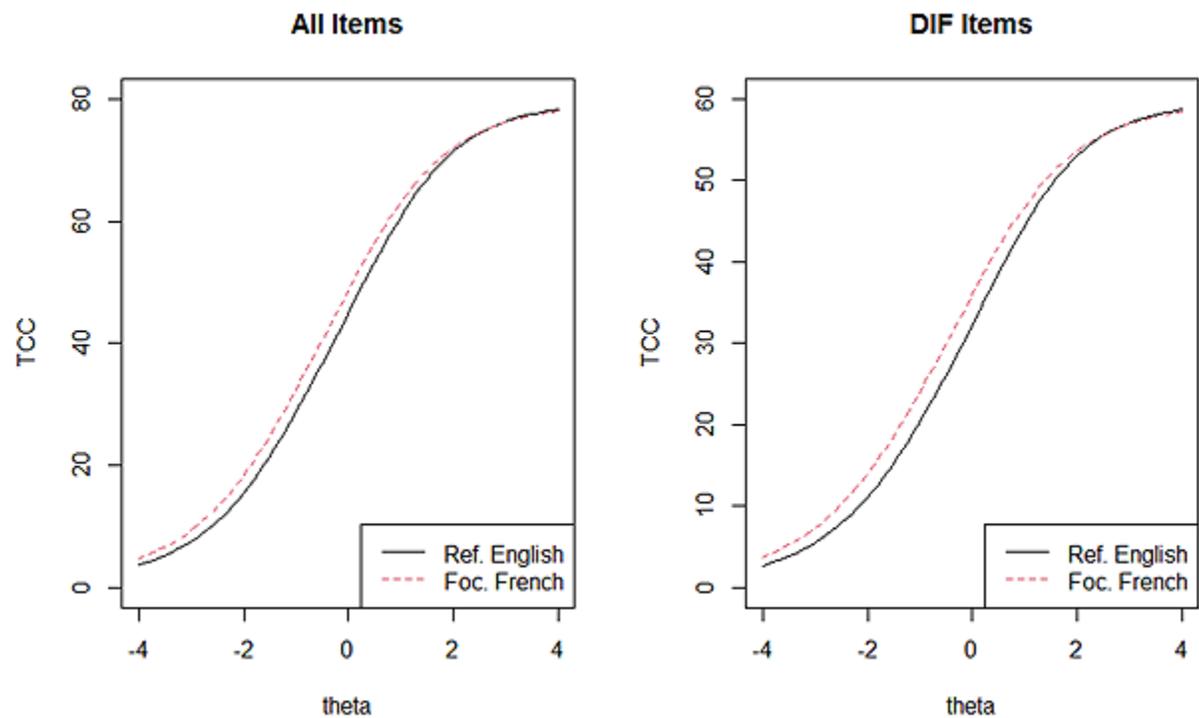


FIGURE 11.3 Détection du FDI selon la langue : courbes caractéristiques du test pour les items de mathématiques



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 2019 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 2019. Par conséquent, tous les items en lecture et en sciences étaient des items d'ancrage. La conception de l'évaluation correspondait à la 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. Puisque aucun effet du mode d'administration n'a été détecté dans l'évaluation du PPCE 2019, l'établissement de liens avec le PPCE 2019 s'est fait au moyen d'un calibrage concomitant avec le PPCE 2016. 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 approche avait l'avantage d'effectuer une utilisation maximale de toutes les données disponibles pour l'estimation des paramètres des items (Martin et coll., 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 2019 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;

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 (CMEC, 2011). La décision d'utiliser 2010 comme année de référence au lieu de 2007 a été prise en raison du changement de la 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.

- appliquer la transformation linéaire de l'étape précédente (étape II) aux distributions des niveaux de rendement de l'évaluation actuelle.

La production des scores dans les trois domaines pour le PPCE 2019 (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 (mathématiques, lecture 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 EAP « espérance *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.
- Des constantes, déterminées par la transformation linéaire, ont ensuite été appliquées à ces scores pondérés pour générer les scores finals qui sont directement comparables aux scores de réussite des précédents PPCE.

Il est important de noter que ce processus s'est également appliqué aux sous-domaines des mathématiques.

Estimations des erreurs-types

L'analyse des données du PPCE 2019 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. La méthode bootstrap est un des nombreux types de techniques d'estimation de la variance se fondant sur la répétition. Lohr (2010) discute en détail des méthodes de répétition. Essentiellement, ces méthodes nécessitent l'utilisation de sous-échantillons répétés, ou répliques, à partir des données, en calculant de nouveau l'estimation de l'enquête pondérée pour chaque réplique, et l'échantillon complet, puis en calculant ensuite la variance comme une fonction des estimations obtenues.

Initialement, 1 000 pondérations de répliques ont été créées à l'étape de l'échantillonnage. À l'étape des analyses, il a été déterminé que 500 suffiraient pour la précision des estimés. De plus, le recours à la totalité des 1 000 pondérations de répliques aurait nécessité trop de ressources informatiques et n'aurait pas produit d'estimations plus précises que celles obtenues avec 500 pondérations de répliques. (Il importe de souligner que les évaluations à grande échelle internationales utilisent moins de 100 pondérations de répliques.) Ainsi, un total de 501 pondérations de répliques bootstrap ont été utilisées (*bsrwgt0*–*bsrwgt500*). La pondération bootstrap d'ordre zéro (*bsrwgt0*), qui est la même que la pondération finale des élèves (*fsu*), n'a pas été utilisée dans le calcul de l'erreur-type d'une estimation ponctuelle. Les pondérations de répliques *bsrwgt1*–*bsrwgt500* ont toutefois été utilisées pour ce calcul. L'inclusion de la pondération *bsrwgt0* est conforme aux pratiques de nombreux programmes d'évaluation à grande échelle qui ont recours à cette structure de pondérations de

répliques. Il importe de signaler que les pondérations *bsrwgt0* et *fsw* sont interchangeables dans le calcul d'une estimation ponctuelle donnée.

Présentation des résultats de l'évaluation du PPCE 2019

Des rapports sommaires sur les scores ont été élaborés au niveau pancanadien, au niveau des provinces, selon la langue du système scolaire et selon le sexe pour chacun des trois domaines évalués. Les variations du rendement au fil du temps ont également été comparées pour chacun des trois domaines. Les résultats ont été fournis sous forme de tableaux et de graphiques selon la présentation utilisée pour les rapports publics du PPCE 2007, 2010, 2013 et 2016. Les erreurs-types ont été calculées au moyen des poids de rééchantillonnage bootstrap. Des tests T ont été réalisés pour toutes les comparaisons effectuées à l'aide des ajustements de Bonferroni, selon le nombre de comparaisons, le cas échéant.

Comme lors des évaluations précédentes, le PPCE 2019 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 en ligne ou sous forme imprimée, selon le mode d'administration du test. En ce qui concerne le *Questionnaire du personnel enseignant* et le *Questionnaire de l'école*, le personnel enseignant et les directions d'école ont rempli leur questionnaire sous forme électronique. L'analyse des données des questionnaires a compris les étapes suivantes :

1. analyse préliminaire;
2. statistiques descriptives;
3. analyse corrélacionnelle;
 - i. corrélation simple
 - ii. modélisation de régression linéaire multiple
4. analyse en composantes principales pour créer des variables calculées lorsque cela était approprié;
5. analyse des items pour les échelles élaborées postulées et empiriques;
6. analyse de comparaison des groupes.

Ces analyses statistiques ont été faites pour chacun des trois questionnaires et ont été présentées pour chaque langue. Le PPCE 2019 est le deuxième cycle dans lequel les pondérations du personnel enseignant ont été incluses; lors des cycles précédents, 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. 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

Les questionnaires du PPCE 2019 comprenaient 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; la réponse était considérée comme non valable si l'élève avait choisi plus d'une réponse (sauf exception indiquée ci-dessous). 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 à choix multiple, où l'élève pouvait cocher plusieurs réponses applicables, lesquelles devaient faire l'objet d'un codage. Des variables numériques ou catégorielles dérivées étaient requises pour les analyses quantitatives de ces items. Ces items recodés incluaient les grades et les diplômes dont les enseignantes et enseignants étaient titulaires, les langues dans lesquelles ceux-ci enseignaient les mathématiques, les types de rétroaction donnée par le personnel enseignant aux élèves des classes de mathématiques, les niveaux scolaires à l'intérieur de l'école et l'auto-identification des élèves en tant qu'Autochtones.

Données manquantes

Il y a eu trois types de données manquantes dans les questionnaires du PPCE 2019 :

- 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 de l'école remplis sur papier sont retournés au CMEC après la capture des données.

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

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.

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 des systèmes scolaires, le cas échéant, pour les variables contextuelles des questionnaires de l'élève, du personnel enseignant et de l'école. Elles comprenaient des distributions en pourcentage pour tous les items utilisant des échelles catégorielles ou de type Likert. Les statistiques descriptives incluaient également la moyenne et l'erreur-type.

Analyse corrélacionnelle

L'analyse corrélacionnelle a consisté à :

- 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. Les résultats de l'analyse corrélacionnelle ont ensuite été utilisés pour éclairer et confirmer les décisions sur l'inclusion ou l'exclusion de différentes variables dans les analyses subséquentes, notamment des analyses de régression linéaire multiple, des analyses en composantes principales, des analyses des items et des indices et des analyses de comparaisons par groupes;
- effectuer une analyse de régression linéaire multiple pour prédire le rendement en mathématiques 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 mathématiques, à partir d'un ensemble de variables se rapportant au personnel enseignant;
- 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 mathématiques, à partir d'un ensemble de variables se rapportant à l'école.

Pour l'ensemble du travail d'analyse corrélacionnelle, il a été 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.

Analyse en composantes principales

Dans les questionnaires, des ensembles ou blocs d'items ont servi à l'examen de caractéristiques ou d'attitudes précises. L'analyse en composantes principales (ACP) a été utilisée avec les items des questionnaires du PPCE 2019 pour réduire la complexité de l'analyse et pour obtenir des mesures plus stables.

En premier lieu, une liste de blocs d'items possibles tirés du *Questionnaire de l'élève*, du *Questionnaire du personnel enseignant* et du *Questionnaire de l'école* pouvant être considérés pour l'ACP a été créée en fonction des rapports contextuels antérieurs et du travail du groupe chargé de l'élaboration des questionnaires du PPCE. Dans l'ensemble, 12 blocs d'items du *Questionnaire de l'élève*, 13 blocs

d'items du *Questionnaire du personnel enseignant* et 6 blocs d'items du *Questionnaire de l'école* ont été signalés pour l'ACP.

Pour chacune des 31 ACP distinctes, une analyse descriptive (fréquence, moyenne, écart-type), pondérée et non pondérée, a d'abord été effectuée pour confirmer l'exactitude et la validité des données. Dans certains cas, les items ont dû être codés de nouveau afin de correspondre à des items similaires pour l'ACP avant que celle-ci ne soit réalisée (p. ex., le code de l'item 23 du *Questionnaire du personnel enseignant* a dû être changé au moyen d'une série de huit variables dichotomiques).

La syntaxe SPSS suivante a ensuite été utilisée dans chaque cas. L'option VARIABLES dresse la liste de toutes les variables de la question, et l'option ANALYSE dresse la liste des variables retenues pour l'analyse, lesquelles pourraient changer à chaque itération en raison des suppressions.

```
FACTOR
/VARIABLES SCHQ18E SCHQ18F SCHQ18G SCHQ18H SCHQ18I SCHQ18J SCHQ18K SCHQ18L
SCHQ18M SCHQ18N SCHQ18O SCHQ18P SCHQ18Q
/MISSING PAIRWISE
/ANALYSIS SCHQ18E SCHQ18F SCHQ18G SCHQ18H SCHQ18I SCHQ18J SCHQ18L SCHQ18M SCHQ18N
SCHQ18O SCHQ18P SCHQ18Q
/PRINT UNIVARIATE INITIAL CORRELATION SIG KMO EXTRACTION ROTATION
/FORMAT BLANK(.30)
/PLOT EIGEN
/CRITERIA MINEIGEN(1) ITERATE(25)
/EXTRACTION PC
/CRITERIA ITERATE(25) DELTA(0)
/ROTATION OBLIMIN
/SAVE REG(ALL)
/METHOD=CORRELATION.
```

Les critères suivants ont ensuite été utilisés pour évaluer le résultat de manière à cerner les variables à supprimer.

- Matrice de corrélation : cette matrice fournit des corrélations simples entre tous les items. Pour éviter la multicolinéarité, les corrélations simples supérieures à 0,8 devraient être signalées, et un des items devrait être enlevé de l'analyse (la décision peut être prise en fonction de critères qualitatifs).
- Test de Kaiser-Meyer-Olkin (KMO) et test de Bartlett : une valeur supérieure au seuil de 0,5 dans le test de KMO et une valeur de signification à un niveau de 0,05 dans le test de Bartlett indiquent que les items conviennent pour l'ACP.
- Qualités de représentation de l'extraction : les variables possédant une faible valeur inférieure à 0,2 devraient être retirées, puisque cela indique que les variables ne sont pas compatibles avec la solution de la composante.
- Valeur propre : une valeur propre minimale de 1 est le seuil pour déterminer le nombre de composantes retenues.

- Variance totale expliquée : si le nombre de composantes ayant des valeurs propres supérieures à 1 compte pour moins de 50 p. 100 de la variance expliquée (certaines personnes suggèrent d'augmenter ce pourcentage à au plus 60 p. 100), il est recommandé de ne pas poursuivre l'ACP.
- Matrice de forme : lorsqu'elle utilise une rotation oblique (Oblimin), la matrice de forme représente l'équation de régression, dans laquelle chaque item normalisé observé est exprimé sous forme de fonction des composantes. Si un item donné a une saturation sur plus d'une composante, il devrait être supprimé si son écart de saturation entre les composantes est inférieur à 0,2 (p. ex., la valeur de saturation entre les composantes serait de 0,05 pour un item avec des saturations de 0,414 et de 0,364 sur deux composantes, auquel cas l'item devrait être supprimé).

Les dernières étapes consistent à vérifier la fiabilité des échelles définies dans l'ACP (voir l'exemple de syntaxe SPSS ci-dessous). L'option VARIABLES dresse la liste de toutes les variables dans la question, et l'option ÉCHELLE n'inclut que les variables retenues pour l'analyse.

```
RELIABILITY
/VARIABLES=STUQ27A STUQ27B STUQ27C STUQ27D STUQ27E STUQ27F STUQ27G
STUQ27H STUQ27I STUQ27J STUQ27K STUQ27L STUQ27M STUQ27N STUQ27O STUQ27P
/SCALE('Number') STUQ27A STUQ27D STUQ27E STUQ27J STUQ27L STUQ27M
/MODEL=ALPHA
/STATISTICS=DESCRIPTIVE SCALE
/SUMMARY=TOTAL.
```

Les critères suivants ont ensuite été utilisés pour évaluer la fiabilité des échelles définies.

- Alpha de Cronbach : la fourchette acceptable pour la statistique du coefficient alpha de Cronbach dépend du nombre d'items, où des valeurs alpha supérieures à 0,7 ou plus sont attendues pour une échelle (bien que cette mesure soit difficile à utiliser comme critère absolu). Néanmoins, il importe de faire preuve de circonspection avec les valeurs alpha inférieures à 0,6.
- Colonne « Cronbach's Alpha if Item Deleted » (alpha de Cronbach si l'item est supprimé) : cette colonne présente la valeur alpha si un item donné est supprimé. Si la valeur pour un item est supérieure à la valeur de l'alpha pour l'échelle, cela signifie que l'item ne contribue pas positivement à l'échelle et, par conséquent, il devrait être considéré de le supprimer.

Il importe de noter que, si un item est enlevé à la suite de l'examen à toute étape énumérée ci-dessus, une nouvelle analyse devrait être générée avec les items restants, en répétant ainsi le même processus d'examen jusqu'à ce que seuls les items convenables soient retenus.

Dans l'ensemble, 10 blocs d'items tirés du *Questionnaire de l'élève* ont été retenus, ce qui a donné lieu à la création de 19 scores de composantes; 12 blocs d'items tirés du *Questionnaire du personnel enseignant* ont été retenus, donnant lieu à 20 scores de composantes; et 6 blocs d'items tirés du *Questionnaire de l'école* ont été retenus, donnant lieu à 12 scores de composantes.

Les scores de composantes ainsi obtenus ont ensuite été échelonnés sur la moyenne de 50 et l'écart-type de 10 pour devenir des *scores d'indice*. Pour faciliter l'interprétation, certains scores d'indice

gradués comme celui pour l'indice des activités parascolaires (1 = « oui », 2 = « non ») ont fait l'objet d'un codage inversé à partir des scores des composantes de façon à ce qu'un score d'indice plus élevé indique la présence des activités parascolaires. Dans d'autres cas, certains des scores de composantes ont été inversés comme sous-produit du processus d'ACP, et les scores d'indice gradués ont fait l'objet d'un codage inversé pour la même raison.

Ces scores d'indice ont ensuite été mis en corrélation avec le rendement en mathématiques. En raison de la grande taille de l'échantillon, il a été établi que la majorité des 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, seules les composantes ayant des corrélations de 0,2 ou plus ont été présentées dans le rapport contextuel du PPCE 2019. Les scores d'indice ont également été divisés en quatre quartiles quasiment égaux, pondérés selon les pondérations de l'élève/du personnel enseignant/de l'école, 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 2019. Une analyse a été effectuée pour chaque questionnaire et les résultats ont été présentés par province, par langue des systèmes scolaires et par sexe. Elle s'est concentrée principalement sur les éléments suivants :

- moyenne et écart-type;
- corrélations entre les items de même qu'avec le rendement des élèves;
- distributions des pourcentages pour tous les items d'un même indice, et rendement correspondant des élèves;
- corrélations et régressions entre scores d'indice et rendements des élèves;
- alpha de Cronbach pour chaque indice (voir la section précédente).

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;
- comparer les moyennes de rendement des élèves pour chaque item des indices dérivés des questionnaires de l'élève, du personnel enseignant et de 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, par quartile;
- comparer les scores d'indice des élèves par sexe et par langue des systèmes scolaires;

- comparer à l'échelle provinciale les indices moyens des élèves par rapport aux moyennes pancanadiennes;
- comparer les moyennes de rendement des élèves aux niveaux du personnel enseignant (c'est-à-dire la classe) et de l'école par rapport aux variables se rapportant au personnel enseignant et à l'école.

Description des ensembles de données

Tous les ensembles de données du PPCE 2019 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 du personnel enseignant et des directions d'école. 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, SAS 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 et un fichier « Format » sont fournis pour l'ensemble de données en format Excel et SAS, respectivement. Pour accéder à l'ensemble de données SAS, il faut préalablement lancer le fichier « Format ».

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 la classe; province et langue de l'élève);
- pondération statistique pour les élèves;
- réponses aux items du *Questionnaire de l'élève*;
- scores d'indice associés aux élèves¹⁷;
- scores de rendement pour tous les domaines et pour les sous-domaines des mathématiques et niveaux de rendement pour les mathématiques.

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 ou enseignant et de l'école; province et langue pour chaque enseignante ou enseignant);
- pondération statistique pour le personnel enseignant;

¹⁷ Les scores d'indice associés aux élèves qui n'ont pas été inclus dans le rapport contextuel sont eux aussi contenus dans l'ensemble de données.

- réponses aux items du *Questionnaire du personnel enseignant*;
- scores d'indice associés au personnel enseignant¹⁸.

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 échantillonnés 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 écoles¹⁹.

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 fournies par le personnel enseignant et les directions d'école. Le nombre de cas figurant dans cet ensemble de données n'équivaut pas au nombre de cas inclus dans l'ensemble de données des élèves, car dans quelques écoles, deux enseignantes ou enseignants ou plus ont fait partie de l'échantillon. Par conséquent, les réponses d'un même élève ont dû être reproduites pour être mises en parallèle avec les réponses des deux enseignantes ou enseignants ou plus associés à cet élève. La pondération statistique pour les élèves et la pondération de réplique bootstrap pour ces élèves sont donc ajustées en conséquence.

Il convient donc de noter que l'ensemble de données fusionné est établi uniquement au niveau de l'élève et que seules la pondération pour les élèves *ajustée* et la pondération de réplique bootstrap *ajustée* doivent être utilisées dans toutes les analyses de l'ensemble de données fusionné. Par conséquent, tout résultat sur les caractéristiques du personnel enseignant ou des écoles obtenu d'un ensemble de données fusionné doit être interprété comme suit : dans l'exemple du niveau d'éducation atteint par le personnel enseignant, tout résultat devrait être présenté comme « le pourcentage d'élèves dont les enseignantes et enseignants possèdent un baccalauréat », et *NON* « le pourcentage d'enseignantes et enseignants qui possèdent un baccalauréat ».

Pour les analyses à l'échelle du personnel enseignant ou des écoles, l'ensemble de données du personnel enseignant ou des écoles doit être utilisé avec la pondération statistique pour le personnel enseignant ou pour l'école, respectivement.

¹⁸ Les scores d'indice associés au personnel enseignant qui n'ont pas été inclus dans le rapport contextuel sont eux aussi contenus dans l'ensemble de données.

¹⁹ Les scores d'indice associés aux écoles qui n'ont pas été inclus dans le rapport contextuel sont eux aussi contenus dans l'ensemble de données.

Consultation de l'ensemble de données pour les recherches

Le PPCE, qui est une évaluation pancanadienne accompagnée de questionnaires contextuels bien structurés, représente 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 ou territorial²⁰. 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 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 et les 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);

²⁰ Aucun territoire n'a participé au PPCE 2019.

- 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 et aux 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 et les 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%c3%a9es_d%e2%80%99%c3%a9valuation_pour_le_milieu_de_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.
- BERTRAND, R., et J-G BLAIS. *Modèles de mesure : l'apport de la théorie des réponses aux items*, Presse de l'Université du Québec, 2004
- 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. « SEM of another flavour: Two new applications of the supplemented EM algorithm », *British Journal of Mathematical and Statistical Psychology*, n° 61, 2008, p. 309-329.
- 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.
- CHOI, S. W., L. E. GIBBONS et P. K. CRANE. « *lordif*: An R package for detecting differential item functioning using iterative hybrid ordinal logistic regression/item response theory and Monte Carlo simulations », *Journal of Statistical Software*, vol. 39, n° 8, 2011, p. 1-30. Sur Internet : <http://www.jstatsoft.org/v39/i08/>
- CHOI, S. W., L. E. GIBBONS et P. K. CRANE. *lordif: Logistic ordinal regression differential item functioning using IRT (0.3-3)*, The R project for statistical computing, 2016a. Sur Internet : <https://cran.r-project.org/web/packages/lordif/index.html>
- CHOI, S. W., L. E. GIBBONS et P. K. CRANE. *Package "lordif" (0.3-3)*, The R project for statistical computing, 2016b. Sur Internet : <https://cran.r-project.org/web/packages/lordif/lordif.pdf>
- CIZEK, G. J., et M. B. BUNCH. *Standard setting: A guide to establishing and evaluating performance standards on tests*, 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.
- CONSEIL DES MINISTRES DE L'ÉDUCATION (CANADA). *PPCE de 2010 – Programme pancanadien d'évaluation*, Toronto, CMEC, 2011.
- CONSEIL DES MINISTRES DE L'ÉDUCATION (CANADA). *PPCE 2019 – Cadre d'évaluation*, Toronto, CMEC, 2020. Sur Internet : <http://cmec.ca/docs/pcap/pcap2019/PCAP-2019-Assessment-Framework-fr.pdf>

- CRANE, P.K., G. VAN BELLE et E.B. LARSON. « Test bias in a cognitive test: Differential item functioning in the CASI », *Statistics in Medicine*, n° 23, 2004, p. 241-256.
- 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, 2008, 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, 2005, p. 149-169.
- 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*, Boston College, Chestnut Hill, MA, TIMSS & PIRLS International Study Center, 2012. Sur Internet : http://timssandpirls.bc.edu/methods/pdf/TP11_Scaling_Achievement.pdf
- 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>
- FU, J., et L. MONFELS. *LRDIF_ES: A SAS macro for logistic regression tests for differential item functioning of dichotomous and polytomous items*, Research Memorandum No. RM-16-17, Educational Testing Service, 2016.
- HIDI, S., et D. BERNDORFF. « Situational interest and learning », dans L. Hoffmann, A. Krapp, K.A. Renninger 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.
- JODOIN, M. G., et M. J. GIERL. « Evaluating Type I error and power rates using an effect size measure with the logistic regression procedure for DIF detection », *Applied Measurement in Education*, n° 14, 2001, p. 329-349.
- JOHNSON, M. S. « Marginal maximum likelihood estimation of item response models in R », *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.
- LAVEAULT, D., et J. GRÉGOIRE. *Introduction aux théories des tests en psychologie et en éducation*, 3^e éd., De Boeck, 2014.
- LOHR, S. *Sampling: Design and analysis* (2^e éd.), Duxbury Press, 2010.
- LORD, F. M. « A broad-range tailored test of verbal ability », *Applied Psychological Measurement*, n° 1, 1977, p. 95-100.
- LORD, F. M. *Application of item response theory to practical testing problems*, Erlbaum, 1980.
- 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.
- MALDONADO, G., et S. GREENLAND. « Simulation study of confounder-selection strategies », *American Journal of Epidemiology*, vol. 138, n° 11, 1993, p. 923-936.
- 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
- MCDONALD, R. P. *Nonlinear factor analysis*, coll. « Psychometric Monographs », n° 15, Richmond, VA, Psychometric Corporation, 1967. Sur Internet : <https://www.psychometricsociety.org/free-articles>
- 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*, l'auteur, 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. Sur Internet : <https://cmec.ca/Publications/Lists/Publications/Attachments/381/PCAP-2016-Public-Report-FR.pdf>
- O'GRADY, K., K. HOUME, E. COSTA, A. ROSTAMIAN et Y. TAO. *PPCE 2019 : Rapport de l'évaluation pancanadienne en mathématiques, en lecture et en sciences*, Toronto, Conseil des ministres de l'Éducation (Canada), 2021. Sur Internet : <https://cmec.ca/Publications/Lists/Publications/Attachments/426/PCAP2019-Public-Report-fr.pdf>
- ORGANISATION DE COOPÉRATION ET DE DÉVELOPPEMENT ÉCONOMIQUES (OCDE). *PISA 2006 – Les compétences en sciences, un atout pour réussir*, Paris, PISA, Éditions OCDE, 2006.

- 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>
- 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.
- STATISTIQUE CANADA. *Anglais, français et langues non officielles parlées à la maison selon la géographie, 2001 à 2016*, tableau 15-10-0009-01, 2020. Sur Internet : https://www150.statcan.gc.ca/t1/tbl1/fr/tv.action?pid=1510000901&request_locale=fr
- VECTOR PSYCHOMETRIC GROUP. *IRTPRO guide*, juin 2020. Sur Internet : <https://vpgcentral.com/wp-content/uploads/2020/06/IRTPROGuide.pdf>
- 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.

1. Introduction

Le Conseil des ministres de l'Éducation (Canada) [CMEC] était responsable de la mise en œuvre de la cinquième édition du Programme pancanadien d'évaluation (PPCE). Le PPCE, qui est conçu pour évaluer les compétences des élèves de 8^e année (2^e secondaire au Québec) dans les domaines de la lecture, des mathématiques et des sciences, a été déployé au printemps 2019.

Le CMEC a demandé à Statistique Canada de concevoir et de mettre en œuvre la stratégie d'échantillonnage des écoles aléatoire pour le PPCE 2019. Cette annexe vise à présenter en détail tous les aspects du plan d'échantillonnage. Plus spécialement, cette annexe couvre des sujets comme : les populations cibles et les populations de l'enquête; les bases d'échantillonnage; la stratification, la taille de l'échantillon et les méthodes de sélection de l'échantillon; ainsi que la méthodologie proposée²¹.

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 les populations cibles et les populations de l'enquête, la base d'échantillonnage, la stratification, la taille de l'échantillon et les méthodes de sélection de l'échantillon. Ces caractéristiques de plan 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 certaines compétences des élèves de 8^e année/2^e secondaire dans les 10 provinces du Canada. Puisqu'il n'existe pas de liste des élèves de 8^e année/2^e secondaire au Canada, l'échantillon du PPCE sera sélectionné au moyen d'un plan d'échantillonnage aléatoire stratifié en deux étapes. Ainsi, l'échantillon d'élèves de 8^e année/2^e secondaire est sélectionné en deux étapes successives.

À la première étape, un échantillon aléatoire d'écoles (les unités d'échantillonnage primaire – UEP) est 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 sont alors énumérées. La deuxième étape de la sélection de l'échantillon consiste à sélectionner aléatoirement une classe de 8^e année/2^e secondaire par école sélectionnée. Les classes de 8^e année/2^e secondaire sont, par conséquent, les unités d'échantillonnage secondaires (UES). Tous les élèves à l'intérieur du champ des classes de 8^e année/2^e secondaire sélectionnées participent à l'évaluation du PPCE.

Pour la première fois, l'évaluation du PPCE a été effectuée en ligne en 2019. Pour évaluer l'incidence du passage d'une évaluation sur papier à une évaluation en ligne, un échantillon supplémentaire d'écoles dans de grandes strates ne faisant pas partie du recensement a été sélectionné. Les élèves de cet échantillon ont fait l'évaluation sur papier.

²¹ Ce document a été préparé par le groupe de consultation statistique du Centre de collaboration internationale et d'innovation en méthodologie (CCIM) de Statistique Canada et par le Conseil des ministres de l'Éducation (Canada).

2.1 Populations cibles et populations de l'enquête

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 valide. Une bonne définition facilite le processus d'échantillonnage et prévient les ambiguïtés. Les populations peuvent être appelées « populations cibles » ou « populations de l'enquête ».

La *population cible* est l'ensemble des éléments au sujet desquels l'information est escomptée et les estimations sont requises. La *population de l'enquête* est un sous-ensemble de la population cible qui est obtenu à partir du plan de l'enquête et d'autres considérations pratiques. La population de l'enquête 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 n'étaient pas financées par les provinces ont été exclues de la population cible. Les types d'élèves suivants ont également été exclus 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 de l'enquête pour cette étude excluait les écoles possédant certaines caractéristiques et les écoles comptant moins de cinq élèves de 8^e année/2^e secondaire. Notre analyse indique que ces exclusions correspondent à au plus 1,7 p. 100.

Deux types d'unités doivent généralement être distingués dans la population de l'enquête : les unités de l'échantillonnage et les unités répondantes. 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 répondantes 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 font référence. Pour la présente étude, la période de référence est définie comme la période pendant laquelle les évaluations du PPCE se déroulent. L'évaluation la plus récente s'est déroulée au printemps 2019.

2.2 Base d'échantillonnage

Le terme *base d'échantillonnage* englobe toute liste, tout matériel ou tout appareil qui délimite, définit et permet l'accès aux éléments de la population de l'enquête. 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 de l'enquête. L'omission et l'inclusion erronées d'unités dans la base entraînent respectivement une sous-couverture et une surcouverture dans l'étude.

Deux bases de l'enquête ont été utilisées pour cette étude – une pour chaque étape de la sélection de l'échantillon. La sous-section 2.2.1 fournit de plus amples détails sur la base d'échantillonnage pour la première étape de la sélection de l'échantillon.

2.2.1 Première étape de la base d'échantillonnage (liste des écoles à l'intérieur du champ)

Pour la première étape de la sélection de l'échantillon, le CMEC a obtenu la liste de toutes les écoles à l'intérieur du champ des provinces. Cette liste a été fournie par Statistique Canada pour utilisation comme première étape de la base d'échantillonnage. Cette liste 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 A.1).

TABLEAU A.1 Variables importantes pour la première étape de la sélection de l'échantillon

| Variable | Description |
|----------------|--|
| ID_ÉCO | Numéro d'identification de l'école |
| PROVINCE | Province dans laquelle l'école est située |
| ID_LANGUE | Langue du conseil ou de la commission scolaire (français ou anglais) |
| ÉLÈVES_8ANNÉE | Nombre d'élèves de 8 ^e année/2 ^e secondaire |
| CLASSES_8ANNÉE | Nombre de classes de 8 ^e année/2 ^e secondaire |
| CODE_EXEMPTION | Indicateur d'exclusion de l'école |

La section 2.1 a défini les quelques types d'écoles qui ne sont pas couverts par cette étude. Pour enlever ces écoles de la base d'échantillonnage avant l'échantillonnage, l'indicateur d'exemption a été utilisé. Le tableau A.2 fournit de la documentation pour les types d'écoles exclues de cette étude.

TABLEAU A.2 Exclusions des écoles par type

| Exemption | Écoles | Élèves | Pourcentage |
|--|--------|--------|-------------|
| Très petite école (moins de cinq élèves) | 633 | 1 423 | 0,4 |
| École pour les élèves ayant des besoins spéciaux | 58 | 1 329 | 0,4 |
| École dans une autre province | 3 | 387 | 0,1 |
| École isolée géographiquement | 13 | 68 | 0,0 |
| École internationale; à l'étranger | 4 | 440 | 0,1 |
| École non financée | 1 | 351 | 0,1 |
| École en milieu hospitalier | 5 | 139 | 0,4 |
| Centre de jeunesse | 27 | 1 525 | 0,0 |
| Commission scolaire à statut particulier | 30 | 750 | 0,2 |
| Total | 774 | 6 412 | 1,7 |

Pour vérifier la couverture de la base du PPCE 2019, Statistique Canada a comparé le nombre d'élèves de 8^e année/2^e secondaire déclaré dans la base avec les projections de la population d'élèves de 13 ans du recensement. Ces chiffres ont aussi été comparés à ceux observés dans les deux bases de cycles précédents du PPCE (tableau A.3). À la lumière de ces résultats, Statistique Canada a vérifié que le nombre d'élèves de 8^e année/2^e secondaire dans la base de 2017 correspondait bien aux chiffres des deux bases de cycles précédents ainsi qu'aux projections du recensement. En 2017, à l'échelle

pancanadienne, Statistique Canada s’est assuré que la base du PPCE offrait une bonne couverture de la population, les chiffres de la base correspondant presque exactement aux projections de la population. À l’échelle provinciale, la couverture de la population se situe en moyenne autour de 95 p. 100. Cependant, ce chiffre ne s’élève qu’à 89,99 p. 100 au Manitoba. Étant donné ces comparaisons, il est possible de conclure que la base du PPCE est de bonne qualité.

TABLEAU A.3 Couverture de la base du PPCE

| | Projections de la population* (élèves de 13 ans) | | | Base** | | | Proportion de la population (%) | | |
|---------------|---|----------------|----------------|----------------|----------------|----------------|---------------------------------|--------------|--------------|
| | 2013 | 2015 | 2017 | 2013 | 2015 | 2017 | 2013 | 2015 | 2017 |
| NL | 5 436 | 5 221 | 5 293 | 5 441 | 5 193 | 5 048 | 100,1 | 99,5 | 95,37 |
| PE | 1 657 | 1 527 | 1 664 | 1 487 | 1 433 | 1 703 | 89,7 | 93,8 | 102,34 |
| NS | 9 785 | 9 146 | 9 046 | 9 792 | 8 674 | 8 818 | 100,1 | 94,8 | 97,48 |
| NB | 8 010 | 7 543 | 7 683 | 7 976 | 7 327 | 7 269 | 99,6 | 97,1 | 94,61 |
| QC | 79 921 | 78 354 | 79 848 | 85 435 | 81 563 | 83 196 | 106,9 | 104,1 | 104,19 |
| ON | 153 091 | 147 322 | 151 152 | 145 756 | 146 765 | 150 318 | 95,2 | 99,6 | 99,45 |
| MB | 16 289 | 15 852 | 16 137 | 14 451 | 14 174 | 14 522 | 88,7 | 89,4 | 89,99 |
| SK | 13 677 | 13 371 | 14 102 | 12 598 | 12 335 | 12 935 | 92,1 | 92,3 | 91,72 |
| AB | 45 386 | 45 354 | 48 118 | 40 094 | 46 578 | 48 599 | 88,3 | 102,7 | 101,00 |
| BC | 47 129 | 45 848 | 46 881 | 46 895 | 45 576 | 45 627 | 99,5 | 99,4 | 97,33 |
| Canada | 380 381 | 369 538 | 379 924 | 369 925 | 369 618 | 378 035 | 97,3 | 100,0 | 99,50 |

* Source : Statistique Canada, Tableau 17-10-0005-01 – 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)

** La base a été produite en fonction de la liste des écoles fournie au CMEC par les provinces.

2.2.2 Deuxième étape de la base d’échantillonnage (liste des classes à l’intérieur du champ)

Une fois qu’un échantillon d’écoles est sélectionné, le CMEC travaille avec les provinces pour énumérer 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 A.4 présente des variables importantes requises pour la deuxième étape de la sélection de l’échantillon.

TABLEAU A.4 Variables importantes pour la deuxième étape de la sélection de l’échantillon

| Variable | Description |
|--|--|
| ID DE LA CLASSE | Au sein de chaque école, un numéro d’identification unique de la classe |
| NOM DE L’ENSEIGNANTE OU ENSEIGNANT | Nom de l’enseignante ou enseignant de mathématiques |
| ÉLÈVES DE 8 ^e ANNÉE/2 ^e SECONDAIRE | Nombre d’élèves de 8 ^e année/2 ^e secondaire dans chaque classe |
| EXEMPTION | Exemption de la classe |

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 indiquées 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 une ou plusieurs des sept accommodations autorisées, doit être exempté. Les élèves ayant des incapacités fonctionnelles, 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 des incapacités intellectuelles ou des problèmes socioaffectifs ou dont les tests ont montré qu'il avait de telles incapacités ou problèmes doit 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.

Toutes les exemptions à l'échelle de la classe ont été approuvées par le CMEC. Les provinces ont aussi dû expliquer pourquoi certaines écoles échantillonnées n'avaient pas participé au PPCE.

2.3 Stratification

La *stratification* est un moyen d'organiser la base d'échantillonnage 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.

Le CMEC a manifesté le besoin de publier des statistiques fiables à l'échelle pancanadienne et provinciale et en fonction de la langue des conseils, commissions et districts scolaires au sein des provinces. Pour garantir un échantillon suffisamment important dans ces volets, les strates du PPCE ont été définies comme la classification croisée de la province selon la langue du conseil ou de la commission scolaire. Dans la province de Québec, en plus d'être stratifiées selon la langue, les écoles ont été stratifiées selon leur type, soit les écoles publiques et les écoles privées.

2.4 Attribution de la taille de l'échantillon

L'*attribution 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 2019, Statistique Canada a entrepris le processus d'attribution de la taille de l'échantillon en considérant la stratégie d'attribution utilisée en 2016. Les provinces ont demandé que la même précision obtenue en 2016 soit maintenue en 2019. Statistique Canada a par

conséquent utilisé le même schéma d'attribution dans toutes les strates, sauf pour la Saskatchewan, secteur anglophone. L'attribution en Saskatchewan, secteur anglophone, a été diminuée à 158 écoles pour s'aligner sur d'autres grandes strates. En plus de la demande du CMEC visant à obtenir la même précision, cette attribution avait aussi l'avantage d'être déjà approuvée par les provinces. Un échantillon supplémentaire de 158 écoles a été sélectionné et les élèves devaient faire l'évaluation sur papier. La taille de l'échantillon a été déterminée de façon que l'erreur type globale n'excède pas 3,5 points. Cet échantillon a été attribué proportionnellement parmi les grandes strates ne faisant pas partie du recensement. Cette façon de faire a été établie par le CMEC pour conserver les strates du recensement comme elles ont été échantillonnées. Puisque l'échantillon de l'évaluation sur papier ne représente pas l'ensemble des provinces, les résultats de l'étude sur le mode d'administration ne s'appliquent qu'aux strates de l'échantillon de l'évaluation sur papier. Le tableau A.5 présente l'attribution de l'échantillon par strate, pour l'échantillon de l'évaluation en ligne et de l'évaluation sur papier.

TABLEAU A.5 Attribution de la taille de l'échantillon des écoles – PPCE 2019

| Strate | Population à l'intérieur du champ | | En ligne | | Sur papier |
|---------------------------------------|-----------------------------------|----------------|--------------|------------------------|------------|
| | Écoles | Élèves | Écoles | Recensement des écoles | Écoles |
| Alberta – anglophone | 756 | 47 226 | 152 | | 20 |
| Alberta – francophone | 18 | 428 | 18 | Oui | |
| Colombie-Britannique – anglophone | 424 | 45 171 | 150 | | 20 |
| Colombie-Britannique – francophone | 13 | 289 | 13 | Oui | |
| Manitoba – anglophone | 320 | 13 753 | 153 | | 8 |
| Manitoba – francophone | 16 | 366 | 16 | Oui | |
| Nouveau-Brunswick – anglophone | 83 | 5 127 | 83 | Oui | |
| Nouveau-Brunswick – francophone | 60 | 2 131 | 60 | Oui | |
| Terre-Neuve-et-Labrador – anglophone | 112 | 4 930 | 112 | Oui | |
| Terre-Neuve-et-Labrador – francophone | 1 | 8 | 1 | Oui | |
| Nouvelle-Écosse – anglophone | 114 | 8 478 | 114 | Oui | |
| Nouvelle-Écosse – francophone | 11 | 339 | 11 | Oui | |
| Ontario – anglophone | 3 105 | 140 943 | 151 | | 62 |
| Ontario – francophone | 176 | 7 414 | 126 | | 8 |
| Ontario – Vide* | 318 | 1 590 | 4 | | |
| Île-du-Prince-Édouard – anglophone | 22 | 1 639 | 22 | Oui | |
| Île-du-Prince-Édouard – francophone | 4 | 64 | 4 | Oui | |
| Québec - anglophone – publiques | 67 | 6 420 | 67 | Oui | |
| Québec - anglophone – privées | 24 | 1 168 | 24 | Oui | |
| Québec - francophone – publiques | 336 | 56 522 | 120 | | 24 |
| Québec - francophone – privées | 115 | 14 992 | 31 | | 8 |
| Saskatchewan – anglophone | 466 | 12 545 | 150 | | 8 |
| Saskatchewan – francophone | 6 | 80 | 6 | Oui | |
| Canada | 6 567 | 371 623 | 1 588 | | 158 |

* Cette strate contient les écoles qui étaient vides dans la base, mais qui avaient la possibilité d'accueillir des élèves de 8^e année en 2019. Toutes les écoles de cette strate appartenaient au secteur anglophone.

Remarque – Bien que les élèves des écoles francophones de l'Île-du-Prince-Édouard et de Terre-Neuve-et-Labrador aient participé à l'échantillon, leurs résultats ont été exclus en raison de leur faible nombre.

2.5 Sélection de l'échantillon des écoles et des classes

La *sélection de l'échantillon* fait référence au processus utilisé pour obtenir l'échantillon de l'enquête à partir de la population de l'enquête. 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 14 des 21 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 était l'échantillonnage systématique (ES), dans le cadre duquel les écoles étaient d'abord 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 attribué en proportion de la taille des écoles. Cette méthode comporte deux avantages. Premièrement, elle est avantageuse sur le plan administratif, puisque des écoles de remplacement sont retenues pour être les écoles directement au-dessus et au-dessous des écoles sélectionnées sur la liste classée des écoles. Le deuxième avantage de cette méthode de sélection de l'échantillon est qu'elle utilise de 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 a déployé tous ses efforts pour confirmer la participation du plus grand nombre possible d'écoles échantillonnées. Cette participation était importante pour limiter le risque de biais de non-réponse. Après avoir communiqué avec toutes les écoles échantillonnées, le CMEC a ensuite communiqué avec les écoles de remplacement pour les écoles échantillonnées qui ne participaient pas. Chaque école échantillonnée ne participant pas a été remplacée, si possible, par la première école de remplacement. La deuxième école de remplacement n'a été utilisée que si l'école de l'échantillon correspondante et la première école de remplacement ne participaient pas. *Si l'école de l'échantillon originale n'était pas admissible, ou qu'il s'agissait d'un type d'école appartenant aux écoles exemptées définies, ou si l'école était fermée, alors aucune école de remplacement n'a été utilisée.* De plus, une école avec peu, mais suffisamment d'élèves de 8^e année/2^e secondaire ne devait pas être remplacée simplement parce que le nombre d'élève était inférieur à celui escompté.

En général, les écoles de remplacement sont utilisées pour contrer la non-réponse (c.-à-d., si et seulement si l'école originale a refusé 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, Statistique Canada ne souhaite 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.

Bref, les écoles de remplacement sont utilisées pour traiter la non-réponse. En 2019, 26 des 1 429 écoles qui ont répondu étaient des écoles de remplacement. Les écoles en dehors du champ ne reçoivent aucun traitement, puisqu'elles représentent le fait que la base n'est pas à jour. Cette taille de l'échantillon perdue est le prix à payer pour avoir une base qui n'est pas à jour.

Pour la deuxième étape de la sélection de l'échantillon (c.-à-d., la sélection des classes), parmi les écoles qui comptaient plus de 20 élèves et qui étaient en mesure d'énumérer 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 comptaient plus de 20 élèves, mais qui n'étaient pas en mesure de dénombrer les classes de 8^e année/2^e secondaire, tous les élèves de ce niveau scolaire ont été dénombrés et un EAS de 20 élèves a été constitué. Dans les écoles où il y avait 20 élèves de 8^e année/2^e secondaire ou moins, tous les élèves ont été recensés. Une fois la liste des classes sélectionnées dressée, le nombre total d'élèves requis dans chaque strate a été comparé au nombre total d'élèves dans les classes échantillonnées. Si le nombre d'élèves échantillonnés dans les classes ne répondait pas aux exigences en matière de nombre requis, telles que définies par Statistique Canada, une deuxième classe a été sélectionnée dans les écoles plus grandes.

2.5.1 Chevauchement minimal avec la TEIMS 2019

Une méthode visant à assurer le chevauchement minimal avec les écoles sélectionnées pour participer à la TEIMS a été utilisée. Cette méthode est décrite dans Chowdhury, Chu et Kaufman²² et permet d'ajuster la probabilité de sélectionner une école pour participer au PPCE 2019 en fonction de sa sélection pour la participation à la TEIMS 2019.

3. Pondérations d'échantillonnage

Les sous-sections 2.5 et 2.5.1 décrivent en détail les méthodes de la première et de la deuxième étape de la sélection de l'échantillon. La présente section vise à préciser les pondérations du plan d'échantillonnage qui découlent des méthodes de sélection de l'échantillon susmentionnées.

Il importe de signaler que la pondération et les rajustements ont été effectués de façon indépendante pour l'échantillon de l'évaluation électronique et celui de l'évaluation sur papier. En d'autres mots, les deux échantillons ont été pondérés de manière à être représentatifs de leur population. L'échantillon de l'évaluation électronique, qui compte 1 588 écoles, offre plus de précision que celui de l'évaluation sur papier, qui ne compte que 158 écoles. La formule suivante s'applique aux deux échantillons. Un ensemble de pondérations a effectivement été déterminé pour ces deux échantillons. L'échantillon de l'évaluation sur papier visait à permettre de déterminer si le passage à une évaluation électronique avait une incidence sur la collecte des données.

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. En considérant que l'ensemble des écoles de la population observée est de taille N , et que h représente les strates desquelles il y a $H = 23$. 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

²² S. Chowdhury, A. Chu et S. Kaufman, « Minimizing overlap in NCES surveys », *Proceedings of the Survey Methods Research Section*, American Statistical Association, 2000, p. 174-179.

8^e année/2^e secondaire dans l'école i . Alors, la pondération du plan d'échantillonnage ($d_i^{écoles}$) pour l'école sélectionnée i est donnée comme :

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

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 :

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

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 :

$$d_i^{élèves} = d_i^{écoles} * raj_i^{classe}$$

En 2019, parmi les 1 429 écoles qui ont répondu, 594 avaient plus de 20 élèves de 8^e année/2^e secondaire et toutes leurs classes ont été dénombrées, et 835 écoles avaient 20 élèves de 8^e année/2^e secondaire ou moins.

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, le CMEC a dû fournir à Statistique Canada le nombre total d'élèves à l'intérieur du champ par classe sélectionnée, ou par école si les classes ne pouvaient pas être dénombrées.

Le CMEC, en collaboration avec les écoles et les provinces, effectue les évaluations du PPCE. En 2019, les mathématiques étaient le domaine principal et la lecture et les sciences étaient les deux domaines secondaires. 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, Statistique Canada a recommandé que la distribution des livrets soit effectuée à partir d'un seul point de départ aléatoire dans chaque province. De plus, il était important 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'était arrêtée, et ainsi de suite. Cette méthode a permis d'assurer la distribution équitable de chaque livret.

3.1 Rajustements de la pondération d'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 2019, parmi les 1 588 écoles, 1 429 ont répondu et 159 n'ont pas répondu.

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 i dans la strate h ($d_{hi}^{école}$, {**pondération_devis_école**})²³ :

$$d_{hi}^{école} = \begin{cases} \frac{\sum_{ieh} X_i}{n_h X_i}, & \text{si } n_h < N_h \\ 1, & \text{si } n_h = N_h \end{cases}$$

Il y a ensuite le facteur de rajustement initial pour la non-réponse pour chaque école de la strate h ($raj_nr_h^{école}$ {**raj_nr_école**}) :

$$raj_adj_h^{école} = \frac{\sum_i I_{hi}^{p-école} + \sum_i I_{hi}^{nr-école}}{\sum_i I_{hi}^{p-école}}$$

où :

$$I_{hi}^{p-école} = \begin{cases} 0 & \text{si l'école ne participe pas} \\ 0 & \text{si l'école est hors-champ} \\ 1 & \text{si l'école participe} \end{cases}$$

$$I_{hi}^{nr-école} = \begin{cases} 1 & \text{si l'école ne participe pas} \\ 0 & \text{si l'école est hors-champ} \\ 0 & \text{si l'école participe} \end{cases}$$

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

$$p_{finale_i}^{école} = raj_adj_h^{école} * d_i^{école}$$

Les rajustements des classes (raj_i^{classe}) ont été calculés pour les écoles participantes. L'étape suivante consistait à calculer les pondérations des élèves. La pondération estimative de chaque élève correspond à la pondération finale de l'école, multipliée par le rajustement de la classe, multiplié par le rajustement de la pondération de l'élève pour tenir compte de la non-réponse.

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-répondants²⁴. 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

²³ Les noms des variables correspondantes dans les dossiers sont présentés entre {} pour l'échantillon de l'évaluation électronique. Pour l'échantillon de l'évaluation sur papier, il faut remplacer « e » par « p ».

²⁴ Le tableau A.7 présente la liste des codes de participation et les codes de réponse associés.

pas participé sont similaires à ceux qui ont participé. Le rajustement pour la non-réponse de l'élève ($raj_nr_i^{élève}$) {**raj_nr_élève**}) pour l'école i dans la classe sélectionnée s'exprime comme suit²⁵ :

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

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, 2 étaient absents et 8 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**}) pour l'élève j de l'école i , dans la strate h , est obtenue de la façon suivante :

$$p_j^{élève} = d_i^{école} * raj_nr_h^{école} * raj_i^{classe} * raj_nr_i^{élève}$$

À l'échelle de l'école, la pondération finale de l'analyse est obtenue de la façon suivante :

$$p_{finale_i}^{école} = raj_nr_h^{école} * d_i^{école}$$

ou en utilisant les noms des variables :

pondération_finale_élève

$$= \text{pondération_devis_école} * \text{raj_nr_école} * \text{raj_classe} * \text{raj_nr_élève}$$

pondération_finale_école

$$= \text{pondération_devis_école} * \text{raj_nr_école}$$

3.2 Pondérations et rajustements à l'échelle de la classe

Les données à l'échelle des classes ont été collectées dans le cadre d'entretiens avec tous les enseignants et enseignantes de mathématiques de chaque classe sélectionnée. Pour certaines classes, plus d'une enseignante ou d'un enseignant a été interrogé. Dans ces cas, il y a eu plus d'une observation par classe, puisque le nombre d'observations par classe correspond au nombre d'enseignantes et enseignants de mathématiques interrogés pour cette classe. Les pondérations à l'échelle de la classe sont égales à la pondération finale de l'école, multipliée par le rajustement de la classe et par le rajustement du nombre d'observations (enseignantes et enseignants). À l'échelle de la classe, la pondération finale de l'analyse est calculée de la façon suivante :

$$p_{finale_i}^{classe} = p_{finale_i}^{école} * raj_i^{classe} * raj_ens_i$$

où :

$$raj_ens_i = \frac{1}{\# \text{ enseignants ayant répondu à l'enquête}}$$

ou en utilisant les noms des variables :

pondération_finale_classe

$$= \text{pondération_finale_école} * \text{raj_classe} * \text{raj_ens}$$

²⁵ Le terme « classe » est défini comme une classe réelle ou un regroupement de (petites) classes pour faire une classe *de facto*.

4. 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 Lohr²⁶. 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 du devis complexe du PPCE, les pondérations bootstrap ont été créées pour les élèves *participants* seulement.

La méthode de Rao et Wu²⁷ pour les estimations des pondérations bootstrap a été utilisée et, étant donné les fractions de sondage importantes dans de nombreuses strates, la variante pour les pondérations bootstrap, telles que présentées dans Beaumont et Patak²⁸ 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, 500 pondérations bootstrap ont été créées (*bsrwt1*,... *bsrwt500*), la pondération bootstrap d'ordre zéro (*bsrwt0*) 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 taux de participation de 100 p. 100, $a_{h,k} = 1$. Ces strates sont présentées dans le tableau A.6.

²⁶ S. Lohr, *Sampling: Design and Analysis*, 2^e éd., Duxbury Press, 2010.

²⁷ J.N.K. Rao et C.E.J. Wu, « Resampling Inference with Complex Survey Data », *Journal of the American Statistical Association*, vol. 83, 1988, p. 231-241.

²⁸ J.F. Beaumont 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.

TABLEAU A.6 Répartition des écoles selon les strates du recensement

| Strate | Nombre d'écoles |
|--|-----------------|
| Colombie-Britannique – francophone | 13 |
| Manitoba – francophone | 16 |
| Nouveau-Brunswick – anglophone | 83 |
| Nouveau-Brunswick – francophone | 60 |
| Terre-Neuve-et-Labrador – francophone | 1 |
| Nouvelle-Écosse – anglophone | 114 |
| Nouvelle-Écosse – francophone | 11 |
| Île-du-Prince-Édouard – anglophone | 22 |
| Île-du-Prince-Édouard – francophone | 4 |
| Saskatchewan – francophone | 6 |
| Écoles qui se sont portées volontaires et ajouts | 3 |

Les écoles qui se sont portées volontaire ont été ajoutées après la création de l'échantillon – il s'agit d'ajouts à l'échantillon initial autres que les écoles de remplacement sélectionnées au préalable. Elles forment leur propre strate, conformément à la recommandation de Kish²⁹ pour les « surprises ».

Tous les élèves avaient un code de réponse. Tous les codes sont présentés dans le tableau A.7 et ont servi à déterminer quels élèves étaient participants.

TABLEAU A.7 Codes de participation

| Code de participation | Description du code | Code de réponse de l'élève |
|-----------------------|--|------------------------------|
| 1 | L'élève est absent. | Non-répondant |
| 2 | L'élève a participé pendant une séance prévue. | Participant |
| 2A | L'élève a participé pendant une séance prévue avec une mesure d'adaptation. | Participant |
| 3 | L'élève a participé pendant une séance de rattrapage. | Participant |
| 4 | L'élève est exempté par l'école. | Exclu |
| 5 | L'élève est exempté parce que des mesures d'adaptation appropriées ne pouvaient pas être offertes. | Exclu |
| 6 | L'élève n'est plus inscrit dans cette école/classe. | A quitté de façon permanente |
| 7 | Les parents ou les élèves ne veulent plus se soumettre à l'évaluation. | Non-répondant |
| 8 | L'élève n'est pas en 8 ^e année/2 ^e secondaire. | Exclu |
| 9 | L'élève reçoit son enseignement à la maison. | Exclu |
| 10 | La feuille de réponse et le livret n'ont pas été remis; seulement les données du questionnaire. | Non-répondant |
| 11 | L'élève a répondu à moins de trois items par domaine et n'a pas rempli au minimum la première section (section 1) du questionnaire contextuel. | Non-répondant |

²⁹ L. Kish, *Survey Sampling*, John Wiley & Sons, 1963.