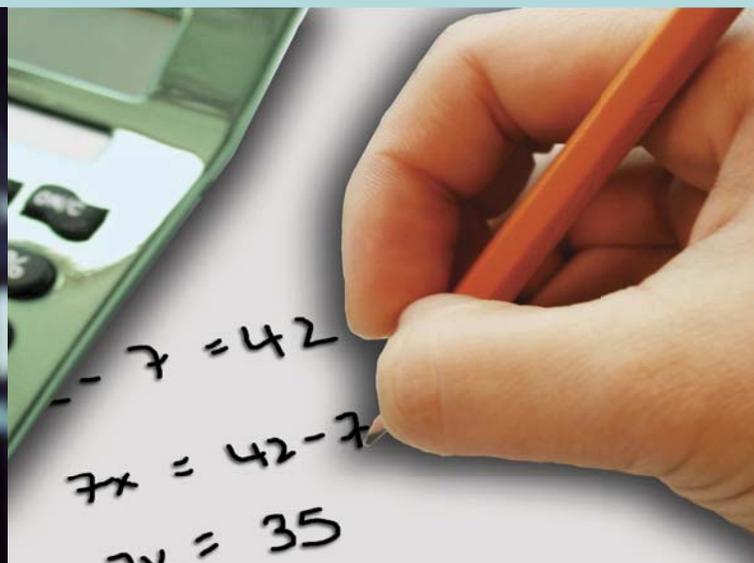


# PPCE 2013

Rapport technique



cmeC

Conseil des  
ministres  
de l'Éducation  
(Canada)

Council of  
Ministers  
of Education,  
Canada

Programme pancanadien d'évaluation

# PPCE 2013

## Rapport technique

### Auteurs

**Kathryn O'Grady**, Conseil des ministres de l'Éducation (Canada)

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



**cme<sup>c</sup>**

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

Par l'entremise du Secrétariat du CMEC, le Conseil agit à titre d'organisation au sein de laquelle les ministères de l'Éducation entreprennent conjointement activités, projets et initiatives, dans des domaines qui intéressent toutes les instances<sup>1</sup>. 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<sup>e</sup> année/2<sup>e</sup> secondaire<sup>2</sup>.*

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](mailto:cmec@cmec.ca)

© 2015 Conseil des ministres de l'Éducation (Canada)

*This report is also available in English.*

---

<sup>1</sup> Dans ce rapport, le terme *instances* englobe les provinces et les territoires qui participent à l'évaluation.

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

## Table des matières

<b>Chapitre 1 – Vue d’ensemble du Programme pancanadien d’évaluation</b> .....	<b>9</b>
Contexte.....	9
Évaluation pancanadienne.....	9
Participation.....	10
Durée de l’évaluation.....	10
Le PPCE dans les deux langues officielles .....	10
<b>Chapitre 2 – Conception et élaboration de l’évaluation</b> .....	<b>12</b>
Conception de l’évaluation .....	12
Conception générale de l’évaluation en sciences.....	13
Forme et type des items .....	13
Items à réponse choisie .....	13
Items à réponse construite.....	14
Items intégrés et contextualisés d’évaluation des attitudes.....	14
Cadre d’évaluation en sciences du PPCE .....	15
Description du domaine.....	15
Organisation du domaine .....	16
Compétences .....	16
Sous-domaines.....	18
Attitudes .....	20
Tableau de spécifications.....	20
Cadre d’évaluation en lecture du PPCE .....	20
Description du domaine.....	21
Organisation du domaine .....	22
Liens entre les évaluations en lecture de 2007, de 2010 et de 2013 .....	22
Cadre d’évaluation en mathématiques du PPCE .....	23
Description du domaine.....	23
Organisation du domaine .....	24
Liens entre les évaluations en mathématiques de 2010 et de 2013.....	25
Groupes de travail.....	25
Élaboration des items .....	26
Traduction et examen des items .....	27
Révision et vérification des items d’évaluation .....	28
Traduction et comparaison des items en français et en anglais .....	28
Lisibilité des contextes et des items du PPCE .....	28
Détermination de la lisibilité.....	29
Résultats des tests de lisibilité.....	30
Révisions linguistiques et typographiques .....	31
Révisions scientifiques .....	32
Révisions psychométriques .....	32
Approbation des items par les instances.....	33

<b>Chapitre 3 – Élaboration des questionnaires contextuels.....</b>	<b>34</b>
Cadre initial pour les questionnaires et principes directeurs.....	34
Questions de base.....	36
Différences entre les sexes.....	36
Répartition et utilisation du temps.....	36
Efficacité et croyances du personnel enseignant en sciences.....	37
Évaluation.....	37
Aménagements (adaptations) et modifications.....	37
Attitudes et motivations.....	37
Stratégies d'apprentissage des élèves.....	38
Stratégies d'enseignement.....	38
Occasions d'apprentissage.....	38
Types d'items.....	39
Questionnaires contextuels.....	39
Questionnaire de l'élève.....	39
Questionnaire du personnel enseignant.....	40
Questionnaire de l'école.....	40
<b>Chapitre 4 – Procédures d'échantillonnage .....</b>	<b>42</b>
Plan d'échantillonnage.....	42
Taille des échantillons.....	44
Utilisation de cahiers d'évaluation multiples.....	44
Présence de plusieurs devis d'échantillonnage.....	45
Choix des écoles.....	47
Bases de données sur les écoles.....	47
Sélection des écoles.....	47
Exclusion d'écoles.....	48
Choix des élèves.....	48
<b>Chapitre 5 – Mise à l'essai des items .....</b>	<b>51</b>
Groupe de travail sur la sélection des items — Mise à l'essai.....	51
Cahiers d'évaluation.....	51
Séance de correction des items.....	51
Capture des données.....	51
Analyse des données.....	52
Groupe de travail sur la sélection des items — Évaluation proprement dite.....	53
Révision du cadre d'évaluation.....	53
<b>Chapitre 6 – Évaluation proprement dite .....</b>	<b>54</b>
Cahiers d'évaluation.....	54
Réexamen du matériel d'évaluation.....	54
Impression des cahiers d'évaluation.....	54
Vérification des documents.....	54
Lettre envoyée aux parents/tuteurs des élèves.....	55

Procédures administratives .....	55
Lieu de l'évaluation .....	55
Déroulement de l'évaluation .....	56
Élèves ayant des besoins spéciaux.....	56
Questionnaires pour la direction de l'école et pour le personnel enseignant .....	57
Participation à l'évaluation et exemptions .....	58
Organisation d'une séance de rattrapage .....	59
Renvoi du matériel d'évaluation.....	59
Séance de correction .....	60
Mise en liasse des cahiers.....	60
Feuilles de correction.....	60
Guide de correction .....	60
Guide de codification .....	61
Chefs de correction .....	61
Chefs de table .....	61
Formation des correctrices et correcteurs .....	61
Fidélité des corrections.....	62
Examens de fidélité.....	62
Fiabilité entre membres du groupe de correction (double correction) .....	64
Fiabilité des tendances .....	65
Correction multiple.....	65
Rapports et commentaires .....	65
Rapport de la coordonnatrice ou du coordonnateur de l'instance.....	66
Rapport de la coordonnatrice ou du coordonnateur de l'école.....	68
Formulaire de commentaires de la correctrice ou du correcteur .....	70
<b>Chapitre 7 – Établir une norme de rendement.....</b>	<b>71</b>
Séances de normalisation .....	71
Sélection d'un groupe de spécialistes.....	71
Versions préliminaires des descripteurs des niveaux de rendement.....	72
Sécurité des documents.....	72
Méthode du signet.....	72
Procédure de normalisation .....	74
Descripteurs des niveaux de rendement .....	76
<b>Chapitre 8 – Traitement des données du PPCE .....</b>	<b>77</b>
Collecte des données .....	77
Saisie des données .....	77
Contrôle de la qualité des données saisies.....	78
Nettoyage des données .....	78
Recodification générale .....	78
Examen des données d'échantillonnage .....	78
Dernier examen des données et préparation de la base de données.....	79

<b>Chapitre 9 – Analyse des données sur le rendement des élèves .....</b>	<b>80</b>
Analyse préliminaire .....	80
Filtrage des données.....	80
Recodification des items.....	80
Données manquantes.....	81
Items non utilisés dans l'évaluation .....	81
Items sans objet.....	82
Items non atteints.....	82
Items omis.....	82
Réponse non valable.....	83
Analyse des items .....	83
Analyse de la théorie classique des items .....	83
Difficulté des items .....	84
Discrimination des items .....	84
Statistiques spécifiques pour les items à CM .....	84
Statistiques spécifiques pour les items à RCS.....	84
Examen pour mettre en évidence les données manquantes.....	84
Fiabilité de l'évaluation du PPCE de 2013 .....	85
Items problématiques.....	85
Analyse de la TRI .....	85
Évaluation de la dimensionnalité du PPCE de 2013 .....	85
Calibrage des items.....	86
Évaluation du caractère convenable des modèles de la TRI .....	87
Fonctionnement différentiel des items .....	87
Liens et équivalences avec les évaluations précédentes pour les domaines secondaires.....	87
Fonctionnement du test .....	88
Production des scores pour le rendement et mise à l'échelle .....	88
Estimations des erreurs-types .....	89
Présentation des résultats de l'évaluation du PPCE de 2013 .....	89
 <b>Chapitre 10 – Analyse des données du questionnaire .....</b>	 <b>90</b>
Analyse préliminaire .....	90
Filtrage des données.....	90
Recodification des items.....	90
Données manquantes .....	91
Statistiques descriptives .....	91
Analyse factorielle.....	91
Analyse des items : analyse de la théorie classique des items.....	92
Analyse de comparaison des groupes .....	92
Analyse corrélacionnelle .....	92
 <b>Chapitre 11 – Base de données du PPCE.....</b>	 <b>94</b>
Description des bases de données .....	94
Base de données des élèves .....	94

Base de données du personnel enseignant .....	94
Base de données des écoles .....	95
Base de données fusionnée — élèves/personnel enseignant/écoles .....	95
Consultation de la base de données pour les recherches .....	95
Conditions et modalités .....	96
Coordonnées .....	97
<b>Bibliographie</b> .....	<b>98</b>

## Liste des tableaux

### Chapitre 2 – Conception et élaboration de l'évaluation

Tableau 2.1 – Nombre de blocs, de scénarios et d'items par domaine et par cahier .....	12
Tableau 2.2 – Répartition des items selon le type et le domaine d'évaluation.....	14
Tableau 2.3 – Pourcentages attribués aux compétences et aux sous-domaines dans l'évaluation scientifique du PPCE de 2013.....	20
Tableau 2.4 – Nombre d'items de lecture par sous-domaine et par cahier .....	22
Tableau 2.5 – Nombre d'items de mathématiques par sous-domaine et par cahier.....	24
Tableau 2.6 – Comparaison de la lisibilité des contextes et des items de l'évaluation du PPCE en sciences.....	30

### Chapitre 4 – Procédures d'échantillonnage

Tableau 4.1 – Nombre d'élèves de 8e année/2e secondaire par population.....	43
Tableau 4.2 – Estimation de la taille d'un échantillon .....	44
Tableau 4.3 – Structure des cahiers d'évaluation .....	44
Tableau 4.4 – Répartition des populations par rapport au premier et au deuxième niveau de l'échantillonnage.....	46
Tableau 4.5 – Paramètres pour la taille des échantillons .....	47

### Chapitre 6 – Évaluation proprement dite

Tableau 6.1 – Résultats des examens de fidélité pour le groupe de correction en sciences n° 1.....	64
Tableau 6.2 – Résultats des examens de fidélité pour le groupe de correction en sciences n° 2.....	64
Tableau 6.3 – Résultats des examens de fidélité pour le groupe de correction en lecture.....	64
Tableau 6.4 – Résultats des examens de fidélité pour le groupe de correction en mathématiques.....	64
Tableau 6.5 – Niveau global d'accord entre membres du groupe de correction pour la double correction .....	65

### Chapitre 7 – Établir une norme de rendement

Tableau 7.1 – Composition des cahiers d'évaluation du PPCE pour les items de sciences.....	73
Tableau 7.2 – Répartition des élèves par niveau de rendement en sciences.....	76

## Chapitre 1 – Vue d’ensemble du Programme pancanadien d’évaluation

### Contexte

Depuis une vingtaine d’années, les ministères de l’Éducation participent à diverses évaluations pour mesurer les compétences des élèves en raisonnement, résolution de problèmes et communication afin de mieux les préparer pour l’avenir. Sur la scène internationale, par l’entremise du Conseil des ministres de l’Éducation (Canada) [CMEC], ils ont participé au Programme international pour le suivi des acquis des élèves (PISA) de 2000, 2003, 2006, 2009 et 2012 (avec plus de 60 pays), au Programme international de recherche en lecture scolaire (PIRLS) de 2011 (avec environ 60 pays) et à l’Enquête internationale sur la maîtrise de l’ordinateur et de l’information (EIMOI) de 2013 (avec environ 20 pays). Certaines instances ont participé individuellement à diverses évaluations sur le rendement des élèves, telles que l’évaluation Tendances de l’enquête internationale sur les mathématiques et les sciences (TEIMS). En outre, la plupart des instances 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 instances ont participé, par l’entremise du CMEC, à l’évaluation *Teacher Education and Development Study – Mathematics* – TEDS-M (étude sur l’éducation et le perfectionnement du personnel enseignant les mathématiques) en 2008 et à l’Enquête internationale sur les enseignants, l’enseignement et l’apprentissage (TALIS) de 2013. L’évaluation du Programme pour l’évaluation internationale des compétences des adultes (PEICA) s’est déroulée en 2012 afin d’étudier de façon générale le niveau des adultes en littératie, en numératie et en résolution de problèmes; cette évaluation a fait intervenir 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 mathématiques, en lecture et écriture et en sciences, effectuées entre 1993 et 2004 par le CMEC. En 2003, les ministres de l’Éducation des provinces et des territoires se sont entendus, par l’entremise du CMEC, pour mettre sur pied le PPCE en vue de remplacer le PIRS.

Les programmes d’études diffèrent d’une instance à une autre dans le pays, de sorte qu’il est difficile d’en comparer les résultats. Toutefois, les jeunes Canadiennes et Canadiens des différentes instances 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 instances, afin que ces dernières disposent de données pancanadiennes comparatives sur les niveaux atteints par les élèves de 8<sup>e</sup> année/2<sup>e</sup> secondaire de tout le pays.

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. Le domaine principal de chaque évaluation du PPCE est l'une de ces disciplines d'apprentissage, mais chaque évaluation comprend des composantes secondaires des deux autres disciplines.

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<sup>e</sup> année/2<sup>e</sup> 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 ou commissions scolaires. Les résultats du PPCE ne sont pas mis à la disposition du personnel enseignant, des conseils et commissions 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 2013, 10 provinces ont participé au PPCE. Même si les Territoires du Nord-Ouest et le Yukon ont, par le passé, participé soit au PIRS soit au PPCE, aucun territoire du Canada n'a participé à cette édition la plus récente du PPCE.

### **Durée de l'évaluation**

Les élèves se sont vu accorder 90 minutes pour répondre aux items de l'évaluation du PPCE. Ils ont eu droit à 30 minutes supplémentaires pour terminer le test, si nécessaire. Il était possible d'accorder du temps supplémentaire aux élèves pour lesquels ce type d'aménagement était prévu dans le programme scolaire ordinaire. Après avoir répondu aux items du cahier d'évaluation, les élèves ont eu 30 minutes pour répondre au questionnaire qui leur était destiné. Les élèves étaient autorisés à utiliser les ressources auxquelles ils avaient normalement accès en cours de langue, de mathématiques et de sciences. Les questionnaires destinés au personnel enseignant et aux écoles ont également servi à se faire une idée plus globale du système éducatif au Canada.

### **Le PPCE dans les deux langues officielles**

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 du PPCE ou par le PPCE proprement dit ont répondu aux mêmes questions, indépendamment de la langue. Les échantillons retenus dans le cadre du PPCE ont été choisis afin qu'ils soient représentatifs à la fois du groupe de langue officielle majoritaire et du groupe de langue officielle minoritaire dans les huit instances comptant un nombre suffisant d'individus pour qu'il soit possible de faire des comparaisons statistiques

valables. En raison de la faible taille des échantillons pour les élèves inscrits dans les écoles francophones de l'Île-du-Prince-Édouard et de Terre-Neuve-et-Labrador, leurs résultats ne figurent pas dans les résultats du PPCE; en revanche, ils ont été inclus dans le calcul des résultats globaux pour ces instances. L'échantillon d'élèves francophones pour la Saskatchewan était certes lui aussi de très petite taille, avec 97 élèves seulement, mais il représentait 85 p. 100 de l'effectif d'élèves francophones de 8<sup>e</sup> année dans la province. Le ministère de l'Éducation de la Saskatchewan a approuvé l'inclusion des données sur cette population dans les rapports sur les résultats.

## Chapitre 2 – Conception et élaboration de l'évaluation

Ce chapitre décrit le processus de conception et d'élaboration du test pour le PPCE de 2013. Il était crucial que l'instrument d'évaluation soit conçu de façon à ce que les items des cahiers fassent une évaluation correcte des compétences des élèves de 8<sup>e</sup> année/2<sup>e</sup> secondaire dans les différentes régions du Canada. Les parties qui suivent présentent les divers stades d'élaboration des cahiers d'évaluation pour les trois disciplines évaluées par le PPCE de 2013, à savoir les sciences, la lecture et les mathématiques. Elles fournissent par conséquent des informations plus détaillées sur la conception et le format du test, les cadres d'évaluation, les groupes de travail, la production des versions initiales des items et la révision de ces versions initiales.

### Conception de l'évaluation

L'évaluation du PPCE est un test papier-crayon qui couvre trois domaines d'évaluation : la lecture, les mathématiques et les sciences. En 2013, le domaine principal était les sciences; et la lecture et les mathématiques étaient les domaines secondaires. Tout comme dans le Programme international pour le suivi des acquis des élèves (PISA), le domaine principal change à chaque édition de l'évaluation; cela signifie que, en 2016, les sciences seront un domaine secondaire et c'est la lecture qui sera le domaine principal du PPCE.

Pour l'évaluation du PPCE, les unités d'évaluation en sciences ont été regroupées en huit groupes ou « blocs » répartis sur quatre cahiers, de sorte que chaque cahier contenait deux blocs d'items en sciences, un bloc d'items en lecture et un bloc d'items en mathématiques. Les quatre cahiers ont été distribués aléatoirement et également aux élèves d'une même classe. Par conséquent, chaque élève a rempli deux des huit blocs d'items d'évaluation en sciences; toutefois, les huit blocs ont été remplis par l'ensemble des élèves d'une classe. De plus, on avait des paires de cahiers comportant des ensembles ou des unités d'items communs, permettant de faire des mesures comparatives du rendement des élèves d'un cahier à l'autre. Tous les cahiers d'évaluation contenaient un questionnaire de l'élève à la fin.

Le tableau 2.1 montre la distribution des blocs, des contextes (ou scénarios) et des items pour les sciences, la lecture et les mathématiques dans l'ensemble des quatre cahiers.

**TABLEAU 2.1 – Nombre de blocs, de scénarios et d'items par domaine et par cahier**

	Sciences			Lecture			Mathématiques		
	Blocs	Scénarios	Items	Blocs	Scénarios	Items	Blocs	Scénarios	Items
<b>Cahier 1</b>	2	10	25	1	2	8	1	2	7
<b>Cahier 2</b>	2	10	24	1	3	8	1	3	8
<b>Cahier 3</b>	2	8	24	1	3	8	1	3	9
<b>Cahier 4</b>	2	9	24	1	3	8	1	2	9

## **Conception générale de l'évaluation en sciences**

Pour l'évaluation du PPCE, le domaine des sciences est divisé en : trois compétences (recherche scientifique, résolution de problèmes et raisonnement scientifique); quatre sous-domaines (nature des sciences, sciences de la vie, sciences physiques et sciences de la Terre); et diverses attitudes, dans un contexte donné. Comme le PPCE en sciences est une évaluation de la culture scientifique, chaque item de l'évaluation est associé à la fois à une compétence et à un sous-domaine. Les items relatifs aux attitudes sont intégrés dans les mises en situation.

Les compétences sont incorporées dans tous les sous-domaines de l'évaluation en sciences, car elles englobent les moyens qu'utilisent les élèves pour répondre aux exigences d'un défi particulier proposé par le test. Le cadre tient compte des programmes d'études actuels de 8<sup>e</sup> année/2<sup>e</sup> secondaire en sciences des instances du Canada, ainsi que des énoncés fondamentaux du *Cadre commun de résultats d'apprentissage en sciences de la nature M à 12 : Protocole pancanadien pour la collaboration en matière de programmes scolaires* (CMEC, 1997)<sup>3</sup>.

Chaque unité d'évaluation présente une mise en situation, suivie d'une série d'items connexes. Les mises en situation choisies pour les unités d'évaluation sont censées susciter l'intérêt des élèves du Canada en 8<sup>e</sup> année/2<sup>e</sup> secondaire et renforcer, par conséquent, leur motivation à l'égard de la participation au test. Les mises en situation sont présentées sous la forme d'un texte initial; celui-ci peut prendre la forme d'un bref récit et inclure des tableaux, des diagrammes et des graphiques. Les personnes chargées de la conception des items ont veillé à ce que les mises en situation soient adaptées au niveau de développement des élèves et ne soient pas propres à une culture ou à un lieu géographique en particulier.

Chaque cahier était composé de huit à 10 unités d'évaluation qui, prises dans leur ensemble, couvraient chacune des compétences et chacun des sous-domaines. Chaque unité comprenait une mise en situation et entre un et six items. Les unités de sciences étaient organisées en huit groupes ou « blocs ».

### **Forme et type des items**

Les responsables de l'élaboration des items du PPCE ont choisi les types d'items qui étaient les plus appropriés pour les questions posées. Il s'agissait d'items à réponse choisie et d'items à réponse construite. Le test comportait environ 75 p. 100 de questions à réponse choisie et 25 p. 100 de questions à réponse construite. Les questions intégrées pour l'évaluation des attitudes correspondaient à 6 p. 100 de l'évaluation.

#### ***Items à réponse choisie***

Les items à réponse choisie (RC) sont des questions qui proposent diverses réponses, parmi lesquelles l'élève en choisit une. Cette catégorie comprend les questions à choix multiple, les questions avec des cases à cocher et les énoncés auxquels il faut répondre par « vrai » ou

---

<sup>3</sup> Pour voir les programmes d'études mis à jour en sciences, veuillez consulter les sites Web officiels des instances.

« faux » ou par « oui » ou « non ». Toutes les questions à choix multiple sont formées d'une amorce suivie de quatre propositions, dont une seulement est la bonne réponse et les trois autres sont des leurres.

### ***Items à réponse construite***

Les items à réponse construite (RCS) sont des questions pour lesquelles l'élève doit fournir une réponse écrite. Cette réponse peut aller d'un syntagme à deux ou trois phrases ou même plusieurs paragraphes, dans le cas des questions plus longues. Il arrive également que l'élève doive créer des tableaux ou des graphiques ou encore dessiner des diagrammes ou concevoir des expériences. Le PPCE en sciences comprenait des questions à réponse construite qui étaient de nature ouverte et qui permettaient de mesurer des habiletés cognitives et des connaissances sur le contenu d'ordre supérieur.

Le recours aux questions à réponse construite constitue également une bonne méthode d'évaluation, puisqu'il fait appel à différentes formes d'évaluation, en fonction de ce que les élèves ont à démontrer. Les questions à réponse construite sont conçues pour permettre l'attribution d'un code à crédit partiel, ce qui constitue un aspect important dans l'évaluation du savoir-faire ou pour les items nécessitant de multiples étapes.

Le tableau 2.2 montre la répartition des différents types d'items en sciences, en lecture et en mathématiques sur l'ensemble des quatre cahiers.

**TABLEAU 2.2 – Répartition des items selon le type et le domaine d'évaluation**

Domaine	Cahier 1		Cahier 2		Cahier 3		Cahier 4	
	RC	RCS	RC	RCS	RC	RCS	RC	RCS
Sciences	18	7	18	6	17	7	17	7
Lecture	6	2	6	2	6	2	6	2
Mathématiques	3	4	5	3	5	4	5	4

### ***Items intégrés et contextualisés d'évaluation des attitudes***

L'acquisition d'attitudes positives est un élément important de l'enseignement et de l'apprentissage des sciences dans la plupart des provinces et des territoires du Canada. Cette importance doit transparaître dans le PPCE en sciences. Celui-ci permet donc de recueillir des données sur les attitudes des élèves au moyen d'items intégrés et contextualisés et du questionnaire de l'élève. Le fait de recueillir des données sur l'attitude des élèves permet de montrer si celles-ci varient selon que l'item est contextualisé (test) ou non (questionnaire) et, le cas échéant, en quoi le rendement s'en trouve affecté. Hidi et Berndoff (1998) estiment que l'intérêt des élèves pour la situation peut avoir un effet puissant sur le fonctionnement cognitif et sur la motivation, mais que son rôle n'a été que rarement et accessoirement étudié. Grâce à l'utilisation combinée d'items contextualisés et du questionnaire de l'élève, le PPCE en sciences pourrait fournir des données qui permettront d'approfondir les recherches dans cette voie.

Le PPCE en sciences comporte suffisamment d'items d'évaluation des attitudes pour préparer une échelle fidèle, mais les réponses aux items liés aux attitudes n'ont pas été incluses dans la notation globale pour la culture scientifique. Néanmoins, ils pourraient fournir un élément important pour ce qui est de brosser le portrait de la culture scientifique des élèves – sujet qui sera exploré dans un numéro à paraître de *L'évaluation... ça compte!* (publication disponible sur le site Web du CMEC).

### **Cadre d'évaluation en sciences du PPCE**

Le *Cadre d'évaluation en sciences du PPCE* a été élaboré par des consultantes et consultants, des coordonnatrices et coordonnateurs en évaluation pancanadienne, des éducatrices et éducateurs spécialisés en sciences et des responsables de l'élaboration des politiques de l'ensemble des instances<sup>4</sup>. Ses orientations sont définies par les objectifs et les résultats des programmes d'études en sciences des instances des populations qui participent à l'évaluation. Le cadre tient également compte de l'état actuel des recherches et des pratiques exemplaires dans l'apprentissage des sciences qui sont conformes aux tendances internationales.

### **Description du domaine**

L'analyse documentaire des programmes d'études en sciences de 8<sup>e</sup> année/2<sup>e</sup> secondaire au Canada menée en préparation du PPCE (CMEC, 2005) indique clairement que la culture scientifique est l'objectif de l'enseignement des sciences dans toutes les instances au Canada. Pour cette évaluation, le *Cadre d'évaluation en sciences du PPCE* définit la culture scientifique comme étant le développement de compétences grâce auxquelles l'élève peut mettre en œuvre des attitudes, des habiletés et des connaissances liées aux sciences et d'une compréhension de la nature des sciences permettant de faire de la recherche, de résoudre des problèmes et de suivre un raisonnement scientifique, afin de comprendre et de prendre des décisions fondées sur des preuves quant aux enjeux liés aux sciences.

La portée de l'évaluation se limite aux concepts et aux habiletés qu'on retrouve et utilise dans les cours de la plupart des élèves de 8<sup>e</sup> année/2<sup>e</sup> secondaire du Canada pendant leurs études. Bien que fondée sur les programmes enseignés aux élèves de 8<sup>e</sup> année/2<sup>e</sup> secondaire, cette évaluation ne couvre pas de manière exhaustive tous les concepts et toutes les habiletés qu'un élève de 8<sup>e</sup> année/2<sup>e</sup> secondaire est censé avoir acquis selon un système en particulier. L'évaluation vise à fournir aux instances les données nécessaires pour orienter leurs politiques en matière d'éducation. Elle n'est pas conçue dans le but de mettre en évidence les points forts ou faibles d'élèves, d'écoles, de districts ou de régions en particulier.

---

<sup>4</sup> Le *Cadre d'évaluation en sciences du PPCE* est disponible à l'adresse <http://www.cmec.ca/docs/pcap/pcap2013/Science-Framework-FR-April2013.pdf>.

## **Organisation du domaine**

### *Compétences*

Il est important que les jeunes comprennent les sciences pour pouvoir participer à la vie en société et pour prendre conscience du fait que les enjeux scientifiques et technologiques ont des répercussions sur leur vie. Les élèves acquièrent une culture scientifique en faisant la démonstration de leurs compétences en recherche scientifique, en résolution de problèmes et en raisonnement scientifique. Le PPCE en sciences met l'accent sur l'évaluation de ces compétences.

*La recherche scientifique fait intervenir la compréhension de la façon dont les recherches sont effectuées en sciences, afin de fournir des explications des phénomènes naturels qui sont fondées sur des preuves.*

La recherche scientifique exige de l'élève qu'il se pose ou formule des questions sur la nature des choses, par l'entremise d'une exploration globale et d'enquêtes ciblées (CMEC, 1997). L'élève se concentre sur les aspects qui expliquent le pourquoi et le comment des sciences.

L'évaluation que fait le PPCE de l'aptitude à utiliser les méthodes scientifiques pour résoudre des problèmes doit montrer que l'élève a les capacités suivantes :

- formuler des hypothèses;
- faire des observations;
- concevoir et mener une enquête;
- organiser et communiquer des informations;
- analyser et interpréter des données (p. ex., au moyen de diagrammes et de tableaux);
- appliquer les résultats d'une enquête scientifique;
- retenir diverses conclusions possibles à la lumière des preuves recueillies;
- motiver ses conclusions en fonction des preuves fournies;
- formuler les hypothèses dont découlent ses conclusions.

*La résolution de problèmes fait intervenir l'utilisation des connaissances et des compétences scientifiques pour résoudre des problèmes dans des contextes sociaux et environnementaux.*

La résolution de problèmes oblige l'élève à chercher à résoudre des problèmes pratiques en trouvant des façons originales de mettre en application ses connaissances scientifiques (CMEC, 1997). L'élève montre qu'il possède cette compétence en utilisant ses connaissances scientifiques, ses habiletés et sa compréhension de la nature des sciences pour résoudre des problèmes scientifiques. Il s'agit d'un aspect de la résolution de problèmes qui consiste à cerner et à présenter le problème, celui-ci étant défini comme une question au regard de laquelle l'élève souhaite atteindre un objectif précis.

L'évaluation que fait le PPCE de l'aptitude à résoudre des problèmes doit montrer que l'élève a les capacités suivantes :

- définir le problème;
- formuler des questions;
- communiquer les objectifs liés au problème;
- résoudre les problèmes en reconnaissant des idées scientifiques;
- choisir des solutions adaptées à un problème donné;
- vérifier et interpréter les résultats (communiquer, réfléchir);
- généraliser les solutions (reconnaître et appliquer les concepts scientifiques dans des contextes habituellement considérés comme n'étant pas scientifiques);
- motiver le choix de la solution et expliquer en quoi celle-ci répond aux critères de résolution du problème;
- formuler les hypothèses dont découlent la solution;
- montrer qu'il est conscient des enjeux du développement durable et de la protection de l'environnement dans le cadre de la résolution du problème.

*Avoir un raisonnement scientifique, c'est raisonner de façon scientifique et établir des liens en appliquant des connaissances et des compétences scientifiques pour prendre des décisions et résoudre des problèmes relatifs aux sciences, à la technologie, à la société et à l'environnement.*

Le raisonnement scientifique exige que l'élève fasse un travail de comparaison, de rationalisation ou de raisonnement à partir d'une théorie ou d'un cadre de référence existant. L'élève montre qu'il a acquis cette compétence en utilisant ses connaissances scientifiques, ses habiletés et sa compréhension de la nature des sciences pour prendre des décisions éclairées basées sur des preuves. Il tire des conclusions ou fait des comparaisons à partir d'une perspective ou d'un cadre de référence existant. L'élève cerne la question ou le problème et cherche à acquérir les connaissances scientifiques qui lui permettront de l'éclairer.

L'évaluation que fait le PPCE de l'aptitude à raisonner de façon scientifique montre que l'élève a les capacités suivantes :

- observer des tendances;
- élaborer des arguments plausibles;
- vérifier des conclusions;
- juger de la pertinence des arguments;
- construire des explications et des arguments pertinents à partir des preuves;
- établir des liens entre les idées scientifiques pour en créer d'autres et produire un tout cohérent;
- raisonner pour prendre une décision éclairée, à partir d'une question donnée, à la lumière des preuves;
- raisonner pour comprendre une question liée aux sciences;
- motiver sa décision en fonction des preuves fournies;

- formuler les hypothèses et les limites qui se rapportent à la décision arrêtée sur une question;
- concevoir et appliquer des modèles;
- faire preuve de respect et de soutien à l'égard des connaissances fondées sur des preuves;
- montrer un intérêt et une sensibilité à l'égard des questions liées aux sciences.

Pour chaque compétence, les élèves sont évalués quant à leur compréhension des pratiques et des processus qui lui sont liés et à leur capacité d'en faire la critique.

### *Sous-domaines*

Les quatre sous-domaines visés par l'évaluation en sciences du PPCE concordent avec les programmes d'études pancanadiens en sciences de toutes les populations participantes et avec les principes de base de la culture scientifique au Canada (CMEC, 1997). Ces quatre sous-domaines sont les suivants : nature des sciences, sciences de la vie, sciences physiques et sciences de la Terre.

#### Nature des sciences

Dans le contexte du PPCE, la nature des sciences désigne la compréhension de la nature des connaissances scientifiques et des méthodes qui favorisent leur évolution. Les sciences proposent une façon de réfléchir au monde biologique et physique et d'apprendre à le connaître, qui est fondée sur l'observation, l'expérimentation et la preuve. Elles se construisent à partir des découvertes du passé. Les théories et connaissances sont continuellement remises en question, modifiées et améliorées, au fur et à mesure que de nouvelles connaissances et théories remplacent les connaissances et théories établies. Le débat scientifique qui accueille de nouvelles observations et hypothèses sert à remettre en question, à communiquer et à évaluer les données grâce à l'interaction entre collègues et à la diffusion des informations par les publications et les présentations. Selon Fensham et Harlen (1999), en développant les capacités qu'ont les élèves d'établir des liens entre preuves et conclusions et de distinguer une opinion d'un énoncé fondé sur des preuves, l'enseignement des sciences favorise une plus grande compréhension des sciences de la part du grand public, ainsi qu'une meilleure compréhension de la prise de décisions fondées sur des preuves, ce qui constitue une composante importante de la culture scientifique.

L'évaluation que fait le PPCE de la compréhension de la nature des sciences montre que l'élève a les capacités suivantes :

- comprendre la relation entre le rassemblement de preuves, l'établissement de liens et les explications proposées dans le développement des connaissances scientifiques;
- faire la distinction entre les procédés et la terminologie qui relèvent des sciences et ceux qui y sont étrangers;
- décrire les procédés de recherche scientifique et de résolution de problèmes propres à la prise de décisions fondées sur des preuves;

- faire la distinction entre données qualitatives et données quantitatives;
- nommer les caractéristiques de la mesure (p. ex., reproductibilité, variation, exactitude ou précision de l'équipement et des méthodes);
- faire la distinction entre divers types d'explications scientifiques (p. ex., hypothèse, théorie, modèle, loi);
- donner des exemples de principes scientifiques qui ont conduit à la création de technologies;
- faire preuve d'une culture scientifique sur les questions liées à la nature des sciences.

Les sous-domaines des sciences de la vie, des sciences physiques et des sciences de la Terre sont évalués au moyen des descripteurs suivants. (Bien que ces descripteurs illustrent les points communs entre les programmes d'études à l'échelle pancanadienne, ils ne constituent pas une liste exhaustive.)

#### Sciences de la vie

- Expliquer et comparer les processus qui maintiennent un organisme en vie.
- Décrire les caractéristiques et les besoins des êtres vivants.
- Distinguer les cellules de leurs composantes.
- Décrire les fonctions des systèmes et l'interdépendance des systèmes relatifs à l'absorption et à la production d'énergie, de nutriments et de déchets.
- Faire preuve d'une culture scientifique sur les questions liées aux sciences de la vie.

#### Sciences physiques

- Décrire les propriétés et les composantes de la matière et expliquer leurs interactions (p. ex., états de la matière [solide, liquide, gaz]; propriétés et transformations de la matière; théorie des particules; masse et volume).
- Faire preuve d'une culture scientifique sur les questions liées aux sciences physiques.

#### Sciences de la Terre

- Expliquer en quoi l'eau est une ressource pour la société.
- Expliquer les tendances quant aux changements et leurs effets sur les ressources en eau sur la Terre (p. ex., répartition de l'eau; climat; usure et érosion; effet de l'eau sur le climat des régions).
- Faire preuve d'une culture scientifique sur les questions liées aux sciences de la Terre.

Bien que la compréhension des relations entre les sciences et la technologie compte pour beaucoup dans l'acquisition d'une culture scientifique, il faut souligner très clairement que le PPCE en sciences n'est pas conçu en vue d'évaluer la culture technologique des élèves qui se soumettent à cette évaluation.

## **Attitudes**

Les attitudes à l'égard des sciences déterminent l'intérêt des élèves pour les professions scientifiques (Osborne, Simon et Collins, 2003). Puisque la création de nouvelles connaissances scientifiques est essentielle à la croissance économique, l'attitude des élèves à l'égard des sciences est un sujet de préoccupation pour les sociétés et fait d'ailleurs l'objet de débats dans bon nombre de pays (OCDE, 2006).

Pour étudier les attitudes des élèves, l'évaluation en sciences du PPCE évalue :

- l'intérêt envers les questions liées aux sciences et la conscience de celles-ci;
- l'adhésion aux connaissances fondées sur des preuves et le respect de ces connaissances;
- la conscience des enjeux du développement durable et de la protection de l'environnement.

## **Tableau de spécifications**

Le tableau de spécifications est un guide d'évaluation qui indique l'importance relative qu'il faut accorder à la mesure de la compréhension des élèves dans des compétences et sous-domaines d'apprentissage divers. Il tient compte du degré de similarité entre les programmes d'études des différentes instances canadiennes. Même si on s'attendait à avoir une proportion plus élevée d'items axés sur la résolution de problèmes dans l'évaluation, bon nombre d'items élaborés pour cette compétence se sont avérés être trop difficiles pour les élèves lors de la mise à l'essai et n'ont donc pas pu être utilisés dans le test proprement dit. Le tableau 2.3 résume les pourcentages attribués à chacune des compétences et à chacun des sous-domaines de l'évaluation du PPCE de 2013.

**Tableau 2.3 – Pourcentages attribués aux compétences et aux sous-domaines dans l'évaluation scientifique du PPCE de 2013**

<b>Compétences</b>		<b>Sous-domaines</b>	
Recherche scientifique	34 %	Nature des sciences	34 %
Résolution de problèmes	12 %	Sciences de la vie	25 %
Raisonnement scientifique	54 %	Sciences physiques	25 %
		Sciences de la Terre	16 %

## **Cadre d'évaluation en lecture du PPCE**

Le cadre d'évaluation en lecture du PPCE de 2013 est le même que celui utilisé pour définir le rendement en lecture lors de l'évaluation de 2007, où la lecture était le domaine principal<sup>5</sup>, ce qui permet des comparaisons au fil du temps entre les trois cohortes.

---

<sup>5</sup> Le *Cadre d'évaluation en lecture du PPCE* est disponible à l'adresse <http://www.cmec.ca/docs/pcap/pcap2007/Reading-Framework-FR-April2013.pdf>.

## ***Description du domaine***

Dans les programmes d'études à l'échelle du Canada, la lecture est considérée comme un processus dynamique et interactif, par lequel la lectrice ou le lecteur construit un sens à partir de textes. L'efficacité du processus dépend de l'interaction entre la lectrice ou le lecteur, le texte, l'intention et le contexte, avant, pendant et après la lecture.

### La lectrice ou le lecteur

Pour dégager le sens d'un texte, les lectrices et lecteurs doivent établir un lien entre le contenu du texte et ce qu'ils savent déjà au moment de l'aborder. Leur expérience personnelle, réelle ou indirecte, détermine dans quelle mesure le contenu et la forme du texte leur sont accessibles. La connaissance de la langue, le degré d'aisance par rapport aux stratégies linguistiques et la connaissance du fonctionnement de la langue à l'écrit influent sur l'idée que l'élève se fait du sens d'un texte.

### Le texte

Les auteures et auteurs produisent des textes à diverses fins, sous diverses formes. À l'heure actuelle, les genres traditionnels sont nombreux à avoir été combinés ou utilisés de façon nouvelle. Les élèves doivent lire un éventail de textes, allant d'œuvres généralement considérées comme des textes de fiction à celles qui sont considérées comme des textes documentaires. Quels qu'ils soient, les textes présentent un niveau plus ou moins complexe sur le plan de la structure, du vocabulaire, de la syntaxe, de l'organisation, des idées, des procédés rhétoriques et du sujet abordé. Pour arriver à bien lire ces formes ou types de textes, il faut que les élèves aient conscience de leur fonctionnement dans différentes situations.

### L'intention de lecture

Le but de la lecture influe sur la façon dont la lectrice ou le lecteur construit le sens du texte. Les élèves lisent des textes à des fins diverses, allant du plaisir qu'ils tirent du contenu et du style du texte aux renseignements pratiques ou aux opinions qu'ils en retiennent. Même si certaines formes ou certains types particuliers de texte sont souvent associés à une intention esthétique ou pragmatique, l'intention de la personne qui les lit peut être différente. Il arrive, par exemple, que l'élève d'un cours de sciences sociales ait à lire un roman pour enrichir ses connaissances sur une culture, une époque, voire un événement particulier.

### Le contexte

Le contexte importe dans tout acte de lecture, puisqu'il influe sur la position de la lectrice ou du lecteur à l'égard de l'imprimé. Le contexte fait précisément référence à l'environnement matériel, affectif, social et institutionnel au moment de la lecture. Le sens que dégage d'un texte la personne qui le lit est le reflet du milieu social et culturel dans lequel elle vit et lit. Les valeurs véhiculées par ses camarades, sa famille et sa communauté ont une incidence sur la position que prend la lectrice ou le lecteur qui aborde le texte.

### **Organisation du domaine**

Étant donné le processus interactif qui met en présence la lectrice ou le lecteur, le texte, l'intention et le contexte, la présente évaluation en lecture porte sur le rapport au texte lu et sur la réaction que suscite ce dernier. Partout au Canada, les programmes d'études de langue font de la compréhension et de l'interprétation, ainsi que de la réaction et de la réflexion, les principaux aspects de l'acquisition de compétences en lecture. Les trois sous-domaines du processus d'intégration de la lecture qui sont évalués portent sur la compréhension, l'interprétation et la réaction au texte (qui comprend la réaction et la réflexion).

*Compréhension* : Les élèves comprennent les informations explicites et implicites que fournit le texte. Ils comprennent en particulier le vocabulaire, les parties du texte, ses éléments et les événements relatés.

*Interprétation* : Les élèves construisent le sens d'un texte en analysant et en synthétisant les parties, les éléments et les événements pour obtenir une perspective plus large ou un sens plus profond. Ils sont capables de cerner le thème ou l'argument et d'étayer leur perception par des renvois à des détails, à des événements, à des symboles, à des schémas et à des caractéristiques du texte.

*Réactions au texte* : Les lectrices et lecteurs réagissent au texte de diverses façons, notamment : en établissant des liens personnels entre certains aspects du texte et les expériences qu'ils ont vécues (directement ou par personne interposée), leurs connaissances, leurs valeurs ou leur point de vue; en ayant une réaction émotive à l'égard des idées centrales ou de certains aspects du texte; ou en évaluant la qualité ou la valeur du texte, éventuellement par rapport à d'autres textes ou facteurs sociaux ou culturels.

Le tableau 2.4 présente la répartition des items de lecture dans les différents cahiers selon les trois sous-domaines.

**TABLEAU 2.4 – Nombre d'items de lecture par sous-domaine et par cahier**

<b>Sous-domaine</b>	<b>Cahier 1</b>	<b>Cahier 2</b>	<b>Cahier 3</b>	<b>Cahier 4</b>
<b>Compréhension</b>	4	3	4	3
<b>Interprétation</b>	3	3	2	3
<b>Réaction au texte</b>	1	2	2	2
<b>Nombre total d'items</b>	8	8	8	8

### **Liens entre les évaluations en lecture de 2007, de 2010 et de 2013**

Le PPCE a pour but de déterminer l'évolution du rendement des élèves au fil du temps. Ce type de comparaison est cependant assez délicat. Le domaine principal du PPCE est défini par rotation dans les trois évaluations du cycle. En raison de cette rotation entre domaine principal et domaines secondaires, les tests en lecture ne sont pas identiques d'une évaluation à la suivante. La lecture était le domaine principal en 2007 et l'évaluation comprenait un grand

nombre d'items, de façon à couvrir de façon étendue les différents sous-domaines décrits dans le *Cadre d'évaluation en lecture du PPCE*. En 2013, tout comme en 2010, la lecture a été un domaine secondaire, avec un nombre limité d'items (environ 20 p. 100). Même si la sélection des items s'est faite à partir de chacun des sous-domaines et avec un éventail de niveaux de difficulté (intervalle pour la valeur  $p$  : de 0,37 à 0,82; intervalle de corrélation : de 0,25 à 0,63) et de niveaux de rendement (de 1 à 3)<sup>6</sup>, l'emploi d'un ensemble plus restreint d'items signifie que ces derniers sont moins représentatifs du cadre d'évaluation dans son ensemble. Pour faciliter les comparaisons entre évaluations, on a élaboré le test en lecture de 2013 à partir d'un sous-ensemble des items de 2007. Ces items, appelés « items d'ancrage », servent à relier les évaluations en lecture de 2007, de 2010 et de 2013 et à mettre en évidence les changements au fil du temps dans le rendement en lecture. Ces items étaient aussi, bien entendu, un sous-ensemble des items du test en lecture de 2010, même s'il y avait cinq items de moins. Il y a eu des items de communs entre les paires de cahiers afin de mieux juger de l'équivalence des cahiers dans le domaine de la lecture.

En 2010, la définition de la population ciblée a changé, pour passer d'une définition selon l'âge (enfants âgés de 13 ans) à une définition selon le niveau scolaire (8<sup>e</sup> année/2<sup>e</sup> secondaire). Comme les résultats étaient rapportés à l'échelle 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 rapportés à l'échelle de l'évaluation de 2007 sur l'échelle de l'évaluation de 2010. Ceci a entraîné des variations des moyennes de 2007 indiquées pour la lecture entre les deux rapports, publiés respectivement en 2007 et en 2010.

### **Cadre d'évaluation en mathématiques du PPCE**

Le cadre d'évaluation en mathématiques du PPCE de 2013 est le même que celui utilisé pour définir le rendement en mathématiques lors de l'évaluation de 2010, où les mathématiques étaient le domaine principal<sup>7</sup>, ce qui permet des comparaisons au fil du temps entre les deux cohortes.

#### ***Description du domaine***

Pour cette évaluation, les mathématiques sont définies de manière générale comme les outils conceptuels que les élèves peuvent utiliser pour renforcer leur capacité de calculer, de décrire et de résoudre des problèmes. Le domaine se divise en quatre sous-domaines et cinq processus. L'évaluation du PPCE porte sur les objectifs d'apprentissage communs en 8<sup>e</sup> année/2<sup>e</sup> secondaire dans toutes les instances participantes au Canada.

---

<sup>6</sup> Veuillez vous référer aux descripteurs des niveaux de rendement pour la lecture dans le rapport public pour le PPCE de 2007, à l'adresse <http://www.cmec.ca/Publications/Lists/Publications/Attachments/124/PPCE2007-Report.fr.pdf>.

<sup>7</sup> Voir le cadre d'évaluation à l'adresse <http://www.cmec.ca/docs/pcap/pcap2013/Math-Framework-April-2013-FR.pdf>.

Quels que soient les termes utilisés pour définir les mathématiques, les programmes d'études au Canada sont structurés de façon à permettre aux élèves [Traduction libre] « d'utiliser les mathématiques dans leur vie personnelle, au travail et lors de la poursuite d'études plus poussées. Tous les élèves ont le droit d'avoir la possibilité de comprendre le pouvoir et la beauté des mathématiques. Les élèves doivent apprendre un nouvel ensemble de principes mathématiques fondamentaux pour calculer aisément et résoudre des problèmes de manière créative et habile » (NCTM, 2000, p. 1).

### **Organisation du domaine**

Le volet de mathématiques du PPCE de 2013 est conforme aux programmes d'études des instances. Le principe dominant de l'évaluation est que l'application des mathématiques est un acte intégré, auquel les habiletés et concepts de divers domaines de contenu sont liés intrinsèquement.

Les sous-domaines de mathématiques du PPCE sont les suivants :

- nombres et opérations (propriétés, représentations équivalentes et ordre de grandeur);
- géométrie et mesures (propriétés des figures à deux dimensions et des formes à trois dimensions, position relative, transformations et mesures);
- régularités et rapports (régularités et expressions algébriques, rapports linéaires et équations); et
- gestion de données et probabilités (collecte et analyse de données, probabilité expérimentale et théorique).

Le tableau 2.5 présente la répartition des items de mathématiques dans les différents cahiers selon les quatre sous-domaines.

**TABLEAU 2.5 – Nombre d'items de mathématiques par sous-domaine et par cahier**

Sous-domaine	Cahier 1	Cahier 2	Cahier 3	Cahier 4
<b>Nombres et opérations</b>	4	0	0	4
<b>Géométrie et mesures</b>	1	5	3	1
<b>Régularités et rapports</b>	0	2	4	3
<b>Gestion de données et probabilités</b>	2	1	2	1
<b>Nombre total d'items</b>	7	8	9	9

Les programmes d'études de mathématiques des diverses instances du Canada s'articulent autour de plusieurs processus mathématiques jugés essentiels à une bonne étude de la matière. Ces processus reflètent la façon dont les élèves acquièrent et appliquent leurs habiletés et connaissances en mathématiques et ne doivent pas être séparés des habiletés et connaissances acquises par l'entremise des programmes d'études. Les cinq processus sont les suivants :

- résolution de problèmes;

- communication;
- représentation;
- raisonnement et preuve;
- établissement de liens.

Les sous-domaines sont des regroupements traditionnels d'habiletés et de connaissances, tandis que les processus se retrouvent dans l'application de tous les sous-domaines.

En ce qui concerne le volet de mathématiques du PPCE de 2013, les contextes des divers scénarios s'appuyaient, autant que possible, sur des situations pertinentes, adéquates et logiques pour des élèves canadiens de 8<sup>e</sup> année/2<sup>e</sup> secondaire.

### ***Liens entre les évaluations en mathématiques de 2010 et de 2013***

Le PPCE a pour but de déterminer l'évolution du rendement des élèves au fil du temps. Ce type de comparaison est cependant assez délicat. Le domaine principal du PPCE est défini par rotation dans les trois évaluations du cycle. En raison de cette rotation entre domaine principal et domaines secondaires, les tests en mathématiques ne sont pas identiques d'une évaluation à la suivante. Les mathématiques étaient le domaine principal en 2010 et l'évaluation comprenait un grand nombre d'items, de façon à couvrir de façon étendue les différents sous-domaines et processus décrits dans le *Cadre d'évaluation en mathématiques du PPCE*. En 2013, les mathématiques ont été un domaine secondaire, avec un nombre limité d'items (environ 20 p. 100). Même si la sélection des items s'est faite à partir de chacun des sous-domaines et avec un éventail de niveaux de difficulté (intervalle pour la valeur  $p$  : de 0,20 à 0,86; intervalle de corrélation : de 0,23 à 0,67) et de niveaux de rendement (de 1 à 4)<sup>8</sup>, l'emploi d'un ensemble plus restreint d'items signifie que ces derniers sont moins représentatifs du cadre d'évaluation dans son ensemble. Pour faciliter les comparaisons entre évaluations, on a élaboré le test en mathématiques de 2013 à partir d'un sous-ensemble des items de 2010. Ces items, appelés « items d'ancrage », servent à relier les évaluations en mathématiques de 2010 et de 2013 et à mettre en évidence les changements au fil du temps dans le rendement en mathématiques.

### **Groupes de travail**

Les groupes de travail se composaient de spécialistes de la lecture, des mathématiques et des sciences en provenance de diverses provinces, qui étaient, pour près de la moitié d'entre eux, bilingues. Ces spécialistes ont travaillé de façon approfondie sur le PPCE et ont participé à diverses étapes du projet, dont l'élaboration du cadre d'évaluation, la préparation des premières versions des items, la validation et la révision des items et la comparaison entre les items en français et les items en anglais. Certains ont également participé aux séances de correction des items lors de la mise à l'essai de l'évaluation et lors de l'évaluation proprement dite.

---

<sup>8</sup> Veuillez vous référer aux descripteurs des niveaux de rendement pour les mathématiques dans le rapport public pour le PPCE de 2010, à l'adresse [http://www.cmec.ca/docs/pcap/pcap2010/Francais/2\\_RapportPublic/RapportPublic\\_ppce2010.pdf](http://www.cmec.ca/docs/pcap/pcap2010/Francais/2_RapportPublic/RapportPublic_ppce2010.pdf).

### ***Élaboration des items***

Les documents devant servir à guider toutes les étapes du processus d'élaboration des items ont été préparés pour la réunion des responsables de l'élaboration des tests à Toronto en septembre 2011. En guise de préparation à la réunion, le cadre d'évaluation en sciences qui avait été élaboré au début du PPCE en 2008 a été révisé afin de tenir compte des changements apportés aux programmes d'études en sciences des différentes instances au Canada.

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

Le processus d'initiation comprenait une vue d'ensemble du cadre d'évaluation en sciences, 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<sup>e</sup> année/2<sup>e</sup> secondaire et de correspondre à l'ensemble des programmes d'études au Canada pour le groupe d'âge concerné. L'élaboration des items s'est faite par petits groupes et s'est déroulée simultanément en anglais et en français. Les séances ont fait intervenir un processus itératif dans lequel de petits groupes travaillaient à l'élaboration d'une unité contenant une série de questions autour d'un stimulus correspondant à la fois à un sous-domaine et à une compétence dans le cadre d'évaluation en sciences. 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. 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 la notation des réponses à chaque question. Chaque guide de codification 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. 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 tous les sous-domaines et toutes les compétences étaient bien couverts.

À la fin de cette séance d'élaboration des items, le groupe de travail a réexaminé et révisé le cadre d'évaluation en sciences 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<sup>e</sup> année/2<sup>e</sup> secondaire au Canada. De petits groupes ont ensuite réexaminé les items de sciences des évaluations antérieures pour s'assurer qu'ils représentaient bien le cadre d'évaluation et avaient bien la bonne classification.

Après la séance d'élaboration des items, il restait encore un certain nombre de sujets dans les sous-domaines qui n'étaient pas représentés. Deux fournisseurs indépendants possédant une grande expérience en élaboration d'items ont été sollicités pour rédiger des items sur des sujets spécifiques, en suivant les mêmes lignes directrices pour l'élaboration des items que celles suivies par le groupe de travail.

Comme les instances n'étaient pas toutes en mesure de participer au processus d'élaboration des items, le CMEC a invité les instances à faire don d'unités d'items susceptibles d'être envisagées pour l'évaluation en sciences. Il fallait que les items correspondent au cadre d'évaluation en sciences et portent sur des sujets communs à l'ensemble des élèves de 8<sup>e</sup> année/2<sup>e</sup> secondaire au Canada. Ces items pouvaient émaner d'évaluations des instances elles-mêmes ou d'évaluations locales des niveaux scolaires venant juste avant ou juste après le niveau 8<sup>e</sup> année/2<sup>e</sup> secondaire, du moment qu'ils étaient accompagnés d'une autorisation indiquant qu'il était possible de les adapter, afin de les rendre compatibles avec l'évaluation du PPCE et qu'aucune restriction relative au droit d'auteur ne s'y appliquait. Ce dispositif a conduit la Colombie-Britannique et l'Alberta à faire don d'items pour l'évaluation.

Une étude a été proposée pour tenter de fournir des données longitudinales en sciences, avec l'établissement d'un lien entre les évaluations du PISA et celles du PPCE. Plusieurs unités du PISA tombées dans le domaine public qui portaient sur la culture scientifique et qui correspondaient au cadre d'évaluation en sciences du PPCE ont été choisies en vue de les inclure dans la mise à l'essai de l'évaluation.

### ***Traduction et examen des items***

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

Un groupe de travail chargé d'examiner les items s'est réuni à Toronto en février 2012. Ce groupe était composé de représentantes et représentants du Manitoba, du Québec, de l'Île-du-Prince-Édouard et de Terre-Neuve-et-Labrador. Le but de la rencontre était d'examiner les items de sciences pour vérifier le contenu, le vocabulaire, la traduction, la conformité aux programmes d'études et l'absence de tout préjugé et aussi pour vérifier la classification des items pour les compétences et les sous-domaines décrits dans le cadre d'évaluation en sciences. Le comité a fait, pour chaque item, l'une des trois recommandations suivantes : a) garder l'item tel quel; b) enlever l'item de la banque d'items pour la mise à l'essai; ou c) garder l'item en y apportant des changements mineurs recommandés par les membres du comité. À l'issue du travail de ce comité, les unités achevées contenaient le texte de stimulus et entre deux et six items. Les unités pouvaient contenir un seul type d'item (à réponse choisie ou à réponse construite) ou un mélange des deux types d'items.

À la fin du travail de ce groupe, le cadre d'évaluation en sciences 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 instances. Dans certains cas, les changements apportés aux documents sur les programmes d'études lors du processus

d'élaboration des items ont conduit à l'élimination d'unités de questions portant sur des sujets qui n'étaient plus communs à l'ensemble des programmes d'études pour les élèves de 8<sup>e</sup> année/2<sup>e</sup> secondaire au Canada. D'autres items ont été éliminés en raison de préjugés 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.

### **Révision et vérification des items d'évaluation**

Avant d'inclure les items dans une évaluation, que ce soit pour la mise à l'essai ou pour l'évaluation du PPCE proprement dite, il était important que ces items soient révisés selon divers points de vue par un groupe de spécialistes. Il fallait prendre autant de soin que possible pour s'assurer que les items étaient raisonnables et fournissaient une évaluation exacte des compétences des élèves de 8<sup>e</sup> année/2<sup>e</sup> secondaire partout au Canada. Lors de la révision, le groupe de spécialistes a pris diverses mesures : traduction et comparaison des items en français et en anglais; révisions linguistiques et typographiques; révisions scientifiques; et révisions psychométriques.

### ***Traduction et comparaison des items en français et en anglais***

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

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

### ***Lisibilité des contextes et des items du PPCE***

Lors de leur réunion de l'automne 2012, les coordonnatrices et coordonnateurs des instances ont mentionné leur inquiétude concernant le fait que les élèves participant au PPCE en français, en particulier ceux qui étaient membres de populations francophones minoritaires, risquaient d'avoir un désavantage dans l'évaluation en sciences du PPCE en raison de la langue. Ils ont recommandé un examen plus attentif des contextes et des items de l'évaluation en sciences, afin de mettre en évidence le risque de préjugés relatifs à la langue susceptibles de favoriser les élèves anglophones.

### ***Détermination de la lisibilité***

Il existe plusieurs modèles basés sur des algorithmes qu'il est possible d'utiliser pour analyser la lisibilité d'un texte. Le modèle Flesch-Kincaid<sup>9</sup> se fonde sur une formule faisant intervenir le nombre de mots, le nombre de phrases et le nombre de syllabes. La formule de Kandel et Moles<sup>10</sup> est une version modifiée adaptée aux textes français. Ces deux modèles évaluent la lisibilité du texte et lui attribuent une note entre 0 et 100, les notes plus élevées dénotant des textes plus faciles à lire. Le modèle Flesch-Kincaid calcule également une formule par niveau scolaire, qui indique le score de lisibilité selon le système des niveaux scolaires des États-Unis.

Des enseignantes et enseignants et des spécialistes des programmes d'études ont été consultés concernant leur utilisation de tels modèles lors de la rédaction de documents et de tests pour les élèves. En règle générale, ces tests sont utilisés comme des lignes directrices à des fins de comparaison, mais traités avec prudence. D'après Benjamin,

[Traduction libre] « si, par le passé, les chercheuses et chercheurs ont conçu des centaines de formules pour estimer la difficulté de textes pour les lectrices et lecteurs, l'utilisation de ces formules est restée controversée pendant des décennies, les critiques découlant dans une large mesure de leur application pour créer de nouveaux tests, ainsi que de leur utilisation d'indicateurs de surface prétendant se substituer aux processus cognitifs complexes qui se déroulent lors de la lecture d'un texte. » (2012, p. 1)

Autrement dit, la lecture est un processus cognitif complexe, de sorte que la lisibilité est une propriété plus complexe qu'un algorithme faisant intervenir le nombre de mots, de syllabes et de phrases. Il peut également y avoir d'autres facteurs importants, comme la familiarité des mots, les répétitions de mots et les termes courants utilisés dans la matière visée. Par conséquent, le jugement professionnel était plus important que les tests de lisibilité, même s'il reste possible d'utiliser ces tests à titre indicatif, au lieu d'en faire le seul et unique critère.

Pour les items de l'évaluation du PPCE en sciences, la formule Flesch-Kincaid par niveau scolaire, qui est une fonction dans le logiciel Microsoft Word, a servi à déterminer le niveau scolaire pour chaque unité. Il a été déterminé que l'éventail pour le niveau scolaire en lecture se situait entre 3,5 et 8,9, la majorité des unités se situant au niveau de lecture de la 8<sup>e</sup> année aux États-Unis. Un programme sur le Web appelé *Readability Index Calculator* (outil de calcul de l'indice de lisibilité) a été utilisé pour calculer la lisibilité parce qu'il permettait de déterminer à la fois la lisibilité en anglais et en français. Pour l'anglais, c'est la formule Flesch-Kincaid qui a été utilisée, tandis que, pour le français, c'est la formule de Kandel et Moles qui a été utilisée.

---

<sup>9</sup> Formule proposée par Klare (1988).

<sup>10</sup> Formule proposée pour le français par Kandel et Moles (1958).

### Résultats des tests de lisibilité

Les résultats de l’outil Readability Index Calculator ont indiqué un niveau scolaire plus élevé et un niveau de lisibilité plus faible que l’outil de Microsoft Word, même si les deux utilisaient le même modèle algorithmique. Ces résultats confirment les inquiétudes des éducatrices et éducateurs concernant l’incohérence de tels modèles algorithmiques, qui simplifient à outrance le processus de lecture. L’outil Readability Index Calculator fait bel et bien une comparaison de la facilité de lecture pour les deux langues, ce qui fournit des informations utiles. À titre d’exemple, la comparaison a révélé que l’inquiétude concernant le fait que les élèves participant à l’évaluation en français auraient un texte plus difficile s’avérait non fondée. Seule une unité était plus facile pour les élèves participant à l’évaluation en anglais et une autre unité était au même niveau de lecture dans les deux langues. Toutes les autres unités étaient plus faciles à lire pour les élèves participant à l’évaluation en français. Par conséquent, selon les tests algorithmiques de lisibilité, le texte des items de l’évaluation du PPCE de 2013 en sciences était d’un niveau plus facile pour la lecture pour les élèves participant à l’évaluation en français que pour ceux participant à l’évaluation en anglais. Le tableau 2.6 présente les résultats de l’analyse de lisibilité.

**TABLEAU 2.6 – Comparaison de la lisibilité des contextes et des items de l’évaluation du PPCE en sciences**

Titre du scénario	MS Word	Readability Index Calculator (Web)	
	Niveau scolaire aux É.-U. (Flesch-Kincaid)	Lisibilité en anglais (Flesch-Kincaid)	Lisibilité en français (Kandel et Moles)
Polluants dans l’air	8,6	42	55
Énergie de substitution et consommation d’énergie	8,1	30	47
Traces qui se trouvent près d’un conteneur à déchets	6,2	43	64
La fourmilière	6,9	29	48
Le baseball	7,5	49	65
Si les arbres étaient camouflés avec une odeur différente, les coléoptères pourraient-ils les trouver?	6,2	61	59
L’achat local	8,8	42	53
Changement climatique	8,3	37	60
Enquête sur une scène de crime	6,7	40	79
Causes de la déforestation de l’Amazonie, 2000-2005	8,5	22	38
La région des lacs expérimentaux	8,4	37	64

El Niño	8,7	36	69
Les joints de dilatation	7,3	41	60
Le quart des espèces de mammifères est menacé d'extinction	9,8	29	62
Les girafes	7,5	39	56
Le réchauffement planétaire	8,8	32	63
Travail par temps chaud	7,2	58	67
Sculpture sur glace	8,2	50	61
Les icebergs	7,0	68	79
L'impact des engrais sur les lacs	5,2	39	60
L'importance de l'eau	6,5	32	54
Les planches à roulettes	8,9	30	47
Solutions	8,3	40	50
Les états de la matière	7,8	34	54
Expérience du lit de ruisseau	4,6	59	71
Le chalet	5,8	51	63
Les caries dentaires	7,2	34	65
Traces dans la neige	3,5	41	41
Les arbres et la forêt	8,1	35	61
Les effets de l'eau	8,8	22	62
Érosion et exposition aux intempéries	6,8	50	70

### ***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 cahier. 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 cahiers, que la mise en page était cohérente, que chaque item commençait par un énoncé suivi d'une question, que le nombre de lignes prévu pour la réponse de l'élève correspondait bien à la longueur de la réponse attendue, 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), etc.

### ***Révisions scientifiques***

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

Plusieurs questions ou scénarios de mathématiques et de sciences comprenaient des tableaux, des diagrammes et des graphiques avec des données. Les réviseuses et réviseurs devaient donc vérifier et garantir l'exactitude des informations. Il était aussi possible que les élèves aient à se référer à un tableau ou à un graphique pour trouver la réponse correcte. On disait aux élèves, dans l'item, 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 cahiers des différentes évaluations du PPCE.

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

### ***Révisions psychométriques***

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

Les spécialistes en sciences ont élaboré un guide de codification avec des descripteurs pour les items à réponse construite. Les guides de codification utilisés pour la lecture et les mathématiques sont restés inchangés par rapport à l'évaluation précédente dans laquelle ces matières étaient le domaine principal (soit l'évaluation de 2007 pour la lecture et l'évaluation de 2010 pour les mathématiques), afin de garantir la cohérence des corrections des items et la possibilité d'analyser les changements dans les résultats au fil du temps. On a attribué divers scores aux réponses des élèves. En mathématiques et en sciences, les scores pouvaient être de 0 ou 1 ou de 0, 1 ou 2. En lecture, les scores allaient de 0 à 3. Chaque score comprenait une description complète et un ou plusieurs exemples tirés des réponses des élèves. Les spécialistes ont donc dû passer en revue tous les critères de correction et s'assurer que les scores définis étaient clairs et précis. Cette étape était très importante, parce que, lors de la séance de correction des items dans les trois matières, les correctrices et correcteurs avaient reçu une formation sur chaque item à corriger. Il fallait qu'ils soient en mesure de faire clairement la distinction entre les différents scores, afin de pouvoir attribuer à chaque réponse d'élève le score 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 d'ouverture.

### **Approbation des items par les instances**

Avant d'inclure les items dans la mise à l'essai sur le terrain, il était important d'obtenir l'approbation par les instances des items choisis. Le CMEC a produit trois cahiers d'évaluation, en français et en anglais, et les a envoyés aux instances pour qu'elles les examinent. Le CMEC a obtenu l'approbation des instances pour pouvoir inclure les scénarios et les items dans la mise à l'essai sur le terrain.

## Chapitre 3 – Élaboration des questionnaires contextuels

### Cadre initial pour les questionnaires et principes directeurs

Lors de la préparation à la conception de la première évaluation du PPCE en 2007, le groupe de travail a commencé par examiner des exemples de questionnaires conçus pour trois programmes d'évaluation à grande échelle, à savoir le PIRS, TEIMS et le PISA. Le groupe a estimé que les questionnaires élaborés pour le PPCE devraient être plus courts et se concentrer sur certains points de façon plus explicite que ceux utilisés pour le PIRS et le PISA. Il était d'avis, en particulier, que le temps dont disposaient les élèves était précieux et qu'il fallait nettement simplifier les questionnaires des élèves. Les membres du groupe se sont également entendus pour dire que, afin de rendre les résultats aussi utiles que possible pour la recherche, il fallait concevoir les questionnaires à partir de certains points de recherche bien précis au lieu d'utiliser l'approche générale des questionnaires utilisés antérieurement.

Ils ont par conséquent adopté les principes suivants lors de la conception des questionnaires :

1. inclure dans les questionnaires des données descriptives de base utiles à la fois à l'élaboration des politiques et à la recherche (statut socioéconomique des élèves, renseignements démographiques sur l'école, qualifications du personnel enseignant, etc.);
2. ne pas chercher à reproduire le PISA, en dehors des questions sur les données descriptives de base;
3. tenter d'explorer un plus petit nombre de domaines, de façon plus approfondie;
4. mettre en évidence les enjeux ayant de la pertinence pour les politiques publiques;
5. exclure les domaines qui se sont avérés improductifs dans le PIRS et le PISA;
6. se concentrer sur le domaine principal dans l'élaboration des questions sur les stratégies d'enseignement et d'apprentissage et les comportements;
7. mettre en évidence un nombre limité de domaines conformes aux orientations du Programme pancanadien de recherche en éducation (PPCRE).

Le groupe de travail s'est attardé sur la limitation imposée par le fait que les données sur l'enseignement et sur l'apprentissage étaient des données de nature transversale et portaient sur le court terme. Le groupe a convenu de tenter de poser des questions conçues de façon à explorer le vécu scolaire des élèves à plus long terme.

Aucun des cadres existants n'est ressorti du lot comme étant, en soi, meilleur que les autres. Tout comme l'évaluation du PPCE elle-même n'est fondée explicitement ni sur le programme d'études ni sur la littératie, il faut une approche plus éclectique pour les questionnaires, qui se fonde sur les priorités définies pour la recherche et sur la nécessité de relier les questionnaires au domaine principal de l'évaluation.

Un groupe de travail s'est réuni à Toronto en janvier 2012 pour élaborer les questionnaires. Ce comité se composait de trois spécialistes externes représentant des établissements universitaires et un organisme qui se consacre à la promotion de la culture scientifique, ainsi

que de quatre représentantes et représentants des instances (Manitoba, secteur anglophone du Nouveau-Brunswick, secteur francophone du Nouveau-Brunswick et Terre-Neuve-et-Labrador). Ces différents membres possédaient une grande expertise sur le contenu des programmes d'études de sciences, sur les recherches en éducation, sur les statistiques et sur l'élaboration de questions pour les questionnaires.

Le but de la réunion était d'élaborer trois questionnaires concis se concentrant sur des questions relatives à l'apprentissage et à l'enseignement des sciences (domaine principal de l'évaluation) et susceptibles de fournir des informations contextuelles importantes pour les instances. En guise de préparation à la mise au point des questions, le comité a examiné les questions de questionnaires scientifiques en provenance de plusieurs sources : évaluations précédentes du PPCE (avec pour domaine principal la lecture et les mathématiques), PIRS, PISA et travaux de recherche. Le groupe de travail a pris pour point de départ l'idée que les questionnaires du PPCE devaient être plus courts et plus ciblés et qu'ils devaient inclure des questions probablement capables de produire des données intéressantes. Les décisions concernant la conservation de questions démographiques et de questions relatives aux origines utilisées dans les évaluations précédentes du PPCE ont été prises en fonction des données issues de l'analyse des questionnaires. L'élimination de questions qui avaient été utilisées lors des deux évaluations précédentes mais n'avaient pas produit d'informations utiles a permis au groupe de produire des documents plus concis. Les questionnaires mis au point par le groupe de travail et portant principalement sur les sciences ont été traduits et révisés par le CMEC, puis envoyés aux coordonnatrices et coordonnateurs des instances pour qu'ils les examinent et apportent des révisions supplémentaires.

L'évaluation du PPCE de 2013 comprenait trois questionnaires : un questionnaire pour les élèves participant à l'évaluation, un questionnaire pour leurs enseignantes et enseignants de sciences de 8<sup>e</sup> année/2<sup>e</sup> secondaire et un questionnaire pour les directions des écoles. Pour les trois questionnaires, la structure générale était dérivée de la synthèse de recherche de Wang, Haertel et Walberg portant sur les facteurs associés aux apprentissages en contexte scolaire (Wang, Haertel et Walberg, 1990, 1993, 1994). Ces questionnaires portaient également sur le besoin particulier de cerner les facteurs associés au rendement en sciences. Les questionnaires 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 sciences, 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 derniers n'ont pas de lien évident avec le rendement dans le domaine principal. L'objectif de cette sélection de thèmes était de fournir des renseignements utiles à la recherche en sciences.

## Questions de base

La section de base des questionnaires comprenait un nombre limité de questions servant à décrire et à comparer ou à tenir compte des variables dans les modèles de recherche. Parmi les sujets traités dans le questionnaire des élèves, on trouvait le sexe des élèves, l'appartenance à un groupe autochtone, le contexte familial, le statut socioéconomique, le fait d'être immigrante ou immigrant, la langue parlée à la maison et la langue d'enseignement. Le questionnaire pour le personnel enseignant portait sur les caractéristiques démographiques et les qualifications du personnel enseignant, les tâches affectées au personnel enseignant et le perfectionnement du personnel enseignant. Le questionnaire destiné à la direction de l'école comportait des questions sur les aspects démographiques et l'administration de l'école, sur le contexte communautaire et sur la composition de l'effectif scolaire. Il est apparu que les questions du PPCE de 2007 sur la langue parlée à la maison ne produisaient pas de données assez détaillées pour justifier un rapport spécial sur le rendement des groupes de langue officielle en milieu minoritaire ou majoritaire, de sorte que cette section a été considérablement développée pour le PPCE de 2010. Cependant, comme elle n'a permis de dégager que peu de tendances, la section a été de nouveau réduite pour l'évaluation de 2013.

## Différences entre les sexes

Les différences de rendement en lecture en faveur des élèves de sexe féminin sont un élément qu'on retrouve régulièrement dans les évaluations à grande échelle, tant au niveau national qu'au niveau international. Les écarts de rendement en sciences et en mathématiques tendent à favoriser les élèves de sexe masculin mais sont beaucoup plus petits qu'en lecture. 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 sciences hors de l'école.

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

## Répartition et utilisation du temps

Le temps a été un aspect important dans certaines autres évaluations. L'incidence du temps sur le rendement a aussi de solides fondements théoriques et empiriques. Le PPCE aimerait trouver des moyens d'améliorer la capacité de mesurer la répartition du temps et les pertes de temps en écartant les variables utilisées jusqu'ici qui présentent une faible variance (par exemple, la durée de l'année scolaire) et en posant des questions un peu plus précises sur la gestion du temps et la motivation des élèves vis-à-vis de l'école. Ces variables sont entre autres les suivantes :

- le temps consacré aux matières;
- la durée des périodes de cours;

- les devoirs imposés et faits;
- le temps perdu (en journées, en périodes de cours, au sein des sessions de cours);
- le temps consacré aux apprentissages hors de l'école;
- l'absentéisme;
- la durée des examens.

### **Efficacité et croyances du personnel enseignant en sciences**

L'efficacité du personnel enseignant est définie comme [*Traduction libre*] « la confiance qu'a l'enseignante ou enseignant dans sa propre capacité de favoriser l'apprentissage des élèves » (Hoy, 2000, p. 1). L'enquête mise au point par Riggs et Enochs (1990) a été incluse dans le questionnaire du personnel enseignant afin d'explorer l'influence de l'efficacité du personnel enseignant sur le rendement des élèves en sciences.

### **Évaluation**

De nombreuses instances ont réagi au rendement inquiétant de certains élèves et de certaines écoles en instaurant un programme provincial ou territorial d'évaluation. La forme de ces programmes varie selon l'instance, tout comme leur degré de maturité. En supposant que l'objectif sous-jacent de cette orientation stratégique soit d'améliorer et non pas seulement de décrire le rendement ni de stabiliser les niveaux actuels, il est tout à fait justifié d'étudier les pratiques d'évaluation des instances et particulièrement l'utilisation faite de ces évaluations. L'objectif est ici d'élargir la portée des questions sur l'évaluation. Certains des aspects à développer sont les suivants : les méthodes d'évaluation; la connaissance qu'ont les enseignantes et enseignants des principes de l'évaluation; et la mesure des différents niveaux de réflexion (connaissance des faits, capacité d'appliquer le savoir, de concevoir des enquêtes scientifiques ou d'évaluer les informations).

### **Aménagements (adaptations) et modifications**

Les questionnaires antérieurs du PPCE comprenaient une série de questions sur certains des enjeux de recherche et de politique publique entourant l'adaptation de l'enseignement et de l'évaluation afin de répondre aux besoins des élèves dans la salle de classe. Le contexte politique général dans ce domaine est la forte tendance, dans la plupart des instances, à intégrer les élèves ayant des besoins spéciaux (sur le plan physique, affectif ou intellectuel) dans la salle de classe ordinaire. Les questions posées lors du dernier PPCE portaient sur l'utilisation de la pédagogie différenciée pour tenir compte de la diversité des styles d'apprentissage et l'offre d'un soutien ou d'une aide aux élèves au sein de la salle de classe.

### **Attitudes et motivations**

Ce domaine est examiné de façon relativement détaillée dans le PISA. On constate régulièrement que les questions et les concepts dans ce domaine sont liés au rendement. Mais il est possible de les simplifier considérablement dans le PPCE. On peut faire des recherches approfondies sur ce domaine à l'aide du PISA et il n'est pas nécessaire de reproduire dans le PPCE ce qui se trouve déjà dans le PISA. L'idée de base ici est que le PPCE ne devrait inclure que le nombre minimum d'items permettant d'utiliser les variables liées à l'attitude et à la

motivation dans la recherche sur les méthodes d'enseignement et d'apprentissage. On a élaboré des items sur les attitudes (d'ordre général et relatives à la matière), les champs d'intérêt et la perception de soi.

### **Stratégies d'apprentissage des élèves**

L'étude des méthodes d'apprentissage mises en œuvre par les élèves est considérée comme l'un des éléments fondamentaux du PPCE. Les questions sur cette facette clé ciblaient les stratégies cognitives et métacognitives auxquelles les élèves ont recours en sciences, c'est-à-dire les stratégies que les élèves utilisent quand ils sont confrontés à des tâches de difficultés et de niveaux divers.

### **Stratégies d'enseignement**

Les questionnaires du PIRS et du PISA contenaient tous de longues listes de méthodes pédagogiques auxquelles les élèves (et les enseignantes et enseignants, dans le cas du PIRS) étaient priés de répondre. On y trouvait des questions générales sur le climat disciplinaire, l'utilisation du temps et les interactions entre élèves et personnel enseignant, ainsi que des questions plus spécifiquement liées aux matières. Ces questions concernaient généralement l'expérience de l'élève ou de l'enseignante ou enseignant dans une classe particulière, au cours de l'année de l'enquête. Étant donné cette portée restreinte, les effets de l'enseignement sont sans doute systématiquement sous-estimés. Plutôt que de reprendre les types d'items qui composaient les questionnaires du PIRS et du PISA, le groupe a tenté de concevoir les questionnaires du PPCE de façon à « remonter plus loin » dans l'expérience passée de l'élève en classe. Ce ne sera sans doute pas facile mais, si les efforts portent leurs fruits, nous comprendrons mieux l'expérience plus générale des élèves à l'école et ses liens avec leur rendement.

Un petit ensemble de questions portait donc sur les perceptions du personnel enseignant et des élèves sur ce qui contribue au rendement en sciences. Des renseignements supplémentaires au sujet des stratégies d'enseignement ont été recueillis en posant aux élèves des questions sur leur assiduité à l'école et sur les pratiques en classe du personnel enseignant (propres à la matière). Parmi les sujets abordés dans les questions, on trouve les suivants :

- la perception qu'a le personnel enseignant de ce qui influe sur le rendement en sciences;
- la perception qu'ont les élèves de leur expérience antérieure des sciences à l'école;
- l'expérience avec les sciences et des questions sur la philosophie pédagogique globale de l'école et sur la démarche relative à l'apprentissage des sciences.

### **Occasions d'apprentissage**

Puisque les occasions d'apprentissage sont souvent considérées comme l'un des meilleurs indices du rendement, on a consacré une petite série de questions aux points suivants :

- les antécédents individuels de l'élève dans l'enseignement des sciences;
- les activités parentales liées aux occasions d'apprentissage.

L'une des caractéristiques prometteuses des résultats du PPCE de 2013 pour les classes de 8<sup>e</sup> année/2<sup>e</sup> secondaire était la possibilité d'établir un lien direct entre le rendement des élèves et les trois questionnaires afin de corrélérer les données (résultats sur le plan du rendement) avec les éléments contextuels pour lesquels ces informations ont été recueillies.

### **Types d'items**

Dans le questionnaire de l'élève, le questionnaire de l'école et le questionnaire du personnel enseignant, la plupart des questions proposaient une série de réponses et les personnes interrogées ne pouvaient généralement cocher que l'une des réponses proposées. Les questionnaires contenaient également plusieurs questions d'opinion visant à mesurer les attitudes et les réactions des personnes interrogées. Ces questions cherchaient à quantifier les attitudes des personnes interrogées à l'aide d'une échelle de Likert. On parle d'« échelle de Likert » pour décrire une échelle ordinaire dans laquelle la personne interrogée classe les réponses dans l'ordre selon leur pertinence. On présentait ainsi aux personnes interrogées une série d'énoncés et elles devaient indiquer dans quelle mesure elles étaient d'accord (« tout à fait d'accord », « pas du tout d'accord », etc.) avec chacun d'entre eux. Les questionnaires contenaient également quelques items pour lesquels les personnes interrogées devaient écrire leur réponse, comme « nombre d'heures ou nombre de jours où... ».

### **Questionnaires contextuels**

#### ***Questionnaire de l'élève***

Après la fin de l'évaluation proprement dite, les élèves avaient 30 minutes pour répondre aux questions du questionnaire de l'élève. La plupart de ces questions se rapportaient aux sciences, puisqu'il s'agissait de la matière principale pour le PPCE de 2013. Environ 32 000 élèves ont répondu au questionnaire.

Le questionnaire de l'élève comprenait quatre sections sur les sujets suivants.

1. Des renseignements personnels, concernant soit les parents ou tuteurs de l'élève soit l'élève lui-même. Ces questions ont permis de recueillir des données démographiques sur l'élève (sexe, statut socioéconomique, statut d'immigrante ou immigrant, etc.).
2. Les attitudes et les motivations de l'élève, qui sont généralement liées au rendement scolaire. Ces questions examinaient les attitudes vis-à-vis de l'école et la valeur attachée par les élèves aux sciences tant dans leur vie personnelle que dans la société.
3. L'expérience de l'élève dans l'apprentissage des sciences à la fois en classe et quand il était plus jeune.
4. La répartition du temps pour l'élève. Elle contenait, par exemple, des questions demandant à l'élève combien de temps il consacrait aux devoirs ou à d'autres activités et des questions sur l'absentéisme.

Les données recueillies grâce au questionnaire de l'élève serviront à faire une comparaison et à mettre en évidence des liens entre les variables étudiées et avec le rendement scolaire de l'élève.

### ***Questionnaire du personnel enseignant***

Il est tout aussi pertinent de rassembler des informations auprès du personnel enseignant sur les variables étudiées grâce à l'évaluation. Le questionnaire du personnel enseignant a été rempli par les enseignantes et enseignants de sciences des élèves de 8<sup>e</sup> année/2<sup>e</sup> secondaire sélectionnés pour le PPCE de 2013.

Le questionnaire du personnel enseignant contenait six sections sur les sujets suivants.

1. Des renseignements personnels sur l'enseignante ou enseignant, comme son sexe, sa formation et son expérience.
2. Le perfectionnement professionnel et posait des questions sur le nombre de journées de perfectionnement professionnel et les activités de perfectionnement professionnel à la fois générales et propres aux sciences.
3. La gestion du temps, entre autres la fréquence des devoirs que l'enseignante ou enseignant donnait à ses élèves et les raisons des pertes de temps en salle de classe.
4. Les méthodes d'évaluation utilisées (types d'évaluations, mode d'attribution des notes, évaluation des différentes façons de penser, prise en compte des divers besoins des élèves, etc.).
5. Les méthodes pédagogiques, par exemple, l'organisation de la salle de classe pour l'enseignement, aux méthodes de différenciation de l'enseignement et à la fréquence et à l'utilisation des stratégies et des activités liées aux sciences.
6. L'efficacité de l'enseignement des sciences, les convictions de l'enseignante ou enseignant et ses attitudes vis-à-vis des sciences, notamment ses raisons pour lesquelles les élèves se débrouillent bien ou moins bien en sciences, sa propre assurance, l'évaluation qu'il fait de son propre travail et ses difficultés.

### ***Questionnaire de l'école***

On a demandé, pour les écoles sélectionnées pour participer au PPCE, à la direction de l'école de remplir le questionnaire de l'école. Ce questionnaire a fourni des informations au niveau de l'école dans le contexte de l'évaluation en sciences du PPCE de 2013 et a permis de faire une analyse et d'établir des liens avec les résultats des élèves ayant participé à l'évaluation.

Le questionnaire de l'école se divisait en cinq sections sur les sujets suivants.

1. Des renseignements généraux. On demandait, par exemple, d'indiquer le nombre d'élèves inscrits dans l'école, les niveaux enseignés dans l'école, le pourcentage d'élèves autochtones et la taille de la communauté où se situait l'école.
2. La gestion du temps. Les questions concernaient le temps consacré à l'enseignement des sciences et le taux d'absentéisme.
3. L'évaluation. On demandait entre autres à la direction de l'école d'indiquer ses pratiques en matière de responsabilisation et les difficultés relatives à l'enseignement des sciences dans l'école.
4. Le climat d'enseignement, y compris des questions sur l'accent mis sur l'enseignement et sur la promotion des sciences. La direction de l'école devait indiquer dans quelle mesure

elle était d'accord avec des énoncés sur le climat d'enseignement en sciences dans son école et indiquer à quelle fréquence se déroulaient différents types d'événements (perfectionnement professionnel, soirées d'information pour les parents, expo-sciences, marques de reconnaissance de la réussite des élèves, etc.).

5. Le contexte pédagogique, notamment la pédagogie différenciée et les difficultés dans l'enseignement.

Le questionnaire du personnel enseignant et le questionnaire de l'école étaient reliés aux résultats des élèves de l'école, mais portaient un numéro d'identification afin de préserver la confidentialité des informations.

## Chapitre 4 – Procédures d'échantillonnage

Au printemps 2013, une troisième évaluation a été réalisée dans le cadre du Programme pancanadien d'évaluation (PPCE). Trois domaines ont été évalués : les sciences, la lecture et les mathématiques – les sciences étant le domaine principal. Quatre cahiers d'évaluation ont été utilisés pour l'ensemble des domaines de l'évaluation, la majorité des items portant sur les sciences. Un seul niveau scolaire (8<sup>e</sup> année/2<sup>e</sup> secondaire) a été évalué. Dix-huit populations ont été associées à la démarche.

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

### Plan d'échantillonnage

Entre 1993 et 2004, le Conseil des ministres de l'Éducation (Canada) [CMEC] a été associé à des évaluations pancanadiennes par l'entremise du Programme d'indicateurs du rendement scolaire (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 cahiers d'évaluation sont utilisés. En 2010 et 2013, 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 privilégie deux axes : le calcul de niveaux de rendement et la compilation de résultats moyens.

Le plan d'échantillonnage a dû être adapté à ce contexte. Comme pour les évaluations du PIRS, une procédure à deux étapes a été utilisée : dans un premier temps, la sélection d'écoles participantes et, dans un deuxième temps, le choix d'une classe de 8<sup>e</sup> année/2<sup>e</sup> secondaire dans les écoles retenues. Compte tenu de la taille des populations visées par l'évaluation, un « recensement » des écoles de certains groupes cibles a pu avoir lieu et des élèves ont pu être sélectionnés de ces écoles. Dans certains cas, il y a eu un recensement des élèves de 8<sup>e</sup> année/2<sup>e</sup> secondaire.

Pour être en mesure de généraliser les statistiques produites pour les élèves faisant partie d'un échantillon, celui-ci doit respecter plusieurs critères<sup>11</sup>. 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. Le tableau 4.1 fournit des statistiques sur les différentes populations canadiennes visées par le PPCE. Ces statistiques proviennent des données fournies au CMEC par les instances pour la mise à l'essai en mai 2012 du mécanisme d'évaluation.

---

<sup>11</sup> Dans le cas du recensement des élèves, il n'y a pas d'inférence statistique; les marges d'erreur n'ont habituellement pas besoin d'être compilées.

**TABLEAU 4.1 – Nombre d’élèves de 8<sup>e</sup> année/2<sup>e</sup> secondaire par population<sup>12</sup>**

Population	Nombre d’écoles	Nombre d’élèves
Colombie-Britannique — anglophone	563	46 632
Colombie-Britannique — francophone	13	263
Alberta — anglophone	823	39 820
Alberta — francophone	23	274
Saskatchewan — anglophone	524	10 803
Saskatchewan — francophone	49	1 795
Manitoba — anglophone	423	14 070
Manitoba — francophone	21	381
Ontario — anglophone	2 756	139 680
Ontario — francophone	173	6 076
Québec — anglophone	124	8 472
Québec — francophone	562	76 963
Nouveau-Brunswick — anglophone	88	5 727
Nouveau-Brunswick — francophone	61	2 249
Nouvelle-Écosse — anglophone	148	9 424
Nouvelle-Écosse — francophone	11	368
Île-du-Prince-Édouard	29	1 487
Terre-Neuve-et-Labrador	154	5 441
<b>Total</b>	<b>6 545</b>	<b>369 925</b>

Le plan d’évaluation du PPCE supposait un déroulement de l’évaluation dans un nombre optimal d’écoles. Voici quelques paramètres reliés au plan d’échantillonnage de l’évaluation proprement dite, qui s’est déroulée en mai 2013.

---

<sup>12</sup> Les statistiques fournies par les instances à l’occasion de la mise à l’essai du PPCE ont permis de constituer ce tableau.

## Taille des échantillons

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

L'utilisation de plusieurs cahiers 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 grand d'élèves a donc été sélectionné afin de garantir une marge d'erreur d'au maximum 3 p. 100, avec un niveau de confiance de 95 p. 100. Le tableau 4.2 indique la formule utilisée pour déterminer la taille d'un échantillon par rapport au calcul des distributions de fréquences.

**TABLEAU 4.2 – Estimation de la taille d'un échantillon**

$n = \frac{Nz^2pq}{Nd^2 + z^2pq}$
<b>N</b> = taille de la population
<b>z</b> = valeur de l'abscisse sur la courbe normale correspondant au niveau de confiance désiré
<b>p</b> = proportion observée dans l'échantillon
<b>q</b> = 1 – p
<b>d</b> = précision désirée, c'est-à-dire marge d'erreur que l'on est prêt à accepter

## Utilisation de cahiers d'évaluation multiples

Au PPCE de 2013, quatre cahiers d'évaluation ont été utilisés. Ils ont été conçus par rapport aux domaines évalués, au nombre d'items par domaine et à la difficulté globale par cahier.

**TABLEAU 4.3 – Structure des cahiers d'évaluation**

Cahier	Sciences	Lecture	Mathématiques	Nombre total d'items
1	24	7	7	38
2	22	8	9	39
3	24	9	8	41
4	24	9	8	41

Au moment du traitement des données, les réponses de chaque cahier ont été compilées. Il a été nécessaire de fusionner les résultats obtenus pour constituer une base de données contenant les résultats de l'ensemble des élèves évalués. Plusieurs valeurs de pondération ont été calculées pour tenir compte de la structure des épreuves utilisées.

L'utilisation de plusieurs cahiers a permis de recueillir plus d'informations touchant différents sous-domaines. Comme les statistiques produites pour ces sous-domaines rendent compte du rendement des élèves par rapport aux résultats moyens, les marges d'erreur associées sont acceptables. Ce n'est pas toujours le cas pour les fréquences concernant le rendement global des sous-groupes d'élèves pour chaque discipline évaluée. Les proportions d'élèves associées à certains regroupements de niveaux de rendement auront probablement, dans certains cas, des marges d'erreur supérieures à 3 p. 100 pour un niveau de confiance de 95 p. 100. Les compilations réalisées pour le questionnaire complémentaire des élèves ont permis d'obtenir des marges d'erreur inférieures à 3 p. 100.

Plusieurs populations ont été évaluées dans le cadre de cette activité. Compte tenu de leur taille, il a fallu considérer un certain nombre de possibilités. En ce qui concerne les écoles sélectionnées (premier niveau de l'échantillonnage), il peut s'agir soit d'un nombre fixe d'écoles ou soit d'un recensement de toutes les écoles faisant partie d'une population cible<sup>13</sup>.

### **Présence de plusieurs devis d'échantillonnage**

Le tableau 4.4 présente la répartition des populations par rapport au premier niveau de l'échantillonnage, c'est-à-dire la sélection (colonne 2) ou non (colonne 3) des écoles associées à l'évaluation; s'il s'agit d'un recensement d'écoles, il peut y avoir en plus un recensement des élèves (colonne 4).

---

<sup>13</sup> Par « population cible », on entend ici les écoles admissibles à la sélection après l'exclusion de celles ne répondant pas aux critères retenus par le CMEC ou les instances concernées. La « population globale », pour sa part, est formée de toutes les écoles accueillant des élèves de 8<sup>e</sup> année/2<sup>e</sup> secondaire.

**TABLEAU 4.4 – Répartition des populations par rapport au premier et au deuxième niveau de l'échantillonnage**

1	2	3	4	5
Population	Échantillonnage au premier niveau	Recensement d'écoles	Recensement d'élèves	Nombre d'élèves à évaluer
Colombie-Britannique — anglophone	X			3 300
Colombie-Britannique — francophone			X	259
Alberta — anglophone	X			3 300
Alberta — francophone			X	262
Saskatchewan — anglophone	X			3 300
Saskatchewan — francophone			X	100 <sup>14</sup>
Manitoba — anglophone	X			3 300
Manitoba — francophone			X	376
Ontario — anglophone	X			3 300
Ontario — francophone	X			2 000
Québec — anglophone		X		2 000
Québec — francophone	X			3 300
Nouveau-Brunswick — anglophone		X		2 000
Nouveau-Brunswick — francophone		X		1 000
Nouvelle-Écosse — anglophone		X		2 500
Nouvelle-Écosse — francophone			X	368
Île-du-Prince-Édouard		X		800
Terre-Neuve-et-Labrador	X			1 500

<sup>14</sup> La confusion concernant la classification des écoles à double programme en Saskatchewan (c'est-à-dire des écoles offrant à la fois un programme de langue anglaise et un programme de langue française) comme relevant du district scolaire anglophone ou du district scolaire francophone a entraîné la sélection d'un échantillon d'élèves anglophones de plus grande taille que ce qui était exigé par le cadre d'échantillonnage.

Le tableau 4.5 présente les paramètres ayant été retenus pour fixer la taille des différents échantillons compte tenu des types de populations décrites.

**TABLEAU 4.5 – Paramètres pour la taille des échantillons**

Types de populations*	Paramètres d'échantillonnage
Populations échantillonnées à deux niveaux (colonne 2)	environ 150 écoles; une classe de 8 <sup>e</sup> année/2 <sup>e</sup> secondaire; un quart des élèves sera évalué à l'aide de l'un des quatre cahiers <sup>15</sup>
Recensement d'écoles (colonne 3)	une classe de 8 <sup>e</sup> année/2 <sup>e</sup> secondaire pour chaque école dans la liste; un quart des élèves sera évalué à l'aide de l'un des quatre cahiers
Recensement d'élèves (colonne 4)	tous les élèves dans la liste; un quart des élèves sera évalué à l'aide de l'un des quatre cahiers

\* Remarque : Les numéros de colonne font référence au tableau 4.4

### Choix des écoles

C'est le Centre des données du CMEC qui a choisi les écoles devant participer au PPCE. Cette sélection s'est faite selon des règles identiques pour toutes, à partir des informations fournies par les instances.

### Bases de données sur les écoles

Dans le but de réaliser les travaux reliés à l'échantillonnage, le CMEC devait disposer de bases de données sur chaque population évaluée. Chaque instance devait utiliser un même fichier préparé par le CMEC afin de dresser la liste des écoles et de préparer les autres informations nécessaires.

### Sélection des écoles

Le CMEC a choisi de centraliser certaines opérations reliées à la sélection des différents échantillons d'écoles retenus pour cette évaluation. Cette stratégie assure une plus grande uniformité des techniques utilisées pour exécuter l'une ou l'autre des opérations requises. Cette façon de faire permet aussi de mieux tenir compte du processus d'échantillonnage.

Pour les populations de plus grande taille, c'est-à-dire celles relevant de la catégorie présentée dans la deuxième colonne du tableau 4.4, les élèves ont été choisis dans 150 écoles différentes environ. Pour ces populations, il s'agissait donc d'un plan d'échantillonnage à deux niveaux (écoles et élèves). Dans ces cas, les écoles ont été choisies aléatoirement en tenant compte de leur taille; les strates disponibles ont été prises en compte au moment de leur sélection. Une même école n'est choisie qu'une fois.

<sup>15</sup> Exceptionnellement, plusieurs classes d'une même école pourront être sélectionnées.

Au moment des compilations statistiques, des pondérations ont été attribuées à chaque élève à l'aide d'une technique identique pour toutes les populations. Cette même technique a aussi été utilisée pour le calcul des marges d'erreur.

### **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 ignorées 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 différents paramètres présents dans les fichiers sur les écoles préparés par les instances. 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 instances 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 instance autre que la province ou le territoire. 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<sup>e</sup> année/2<sup>e</sup> secondaire.

### **Choix 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<sup>e</sup> année/2<sup>e</sup> secondaire fréquentant une école sélectionnée à l'évaluation du PPCE. Il a donc fallu échantillonner 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<sup>e</sup> année/2<sup>e</sup> 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 instances a présenté la liste de toutes les écoles admissibles avec des élèves de 8<sup>e</sup> année/2<sup>e</sup> secondaire qui relevaient de l'autorité de son instance.
2. Le CMEC a sélectionné les écoles devant participer au PPCE et envoyé la liste des écoles aux coordonnatrices et coordonnateurs des provinces et des territoires.
3. Les coordonnatrices et coordonnateurs ont communiqué avec les écoles sélectionnées et demandé la liste des classes de 8<sup>e</sup> année/2<sup>e</sup> 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 et des territoires. 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 et des territoires pouvaient choisir de retirer telle ou telle classe de la participation au PPCE. Dans ce cas, il leur fallait communiquer au CMEC pour assurer la sélection d'une classe de substitution. Il fallait que les coordonnatrices et coordonnateurs des instances 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<sup>e</sup> année/2<sup>e</sup> 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 des écoles de dresser la liste de tous les élèves de 8<sup>e</sup> année/2<sup>e</sup> secondaire comme suit :
  - i. quand c'était possible, la liste de tous les groupements de classe de 8<sup>e</sup> année/2<sup>e</sup> secondaire (par exemple, 8 A, 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<sup>e</sup> année/2<sup>e</sup> secondaire, par ordre alphabétique.
6. Après l'évaluation, les instances ont communiqué au CMEC la liste des élèves ayant participé au PPCE de 2013. 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.

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 de l'opération. 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 ou territoriaux. 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.

## **Chapitre 5 – Mise à l’essai des items**

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 des items s’effectue sur un plus grand nombre d’items que l’évaluation proprement dite, afin que seuls les meilleurs soient retenus pour évaluer le rendement des élèves de 8<sup>e</sup> année/2<sup>e</sup> secondaire.

### **Groupe de travail sur la sélection des items — Mise à l’essai**

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

### **Cahiers d’évaluation**

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

### **Séance de correction des items**

La séance de correction s’est déroulée sur quatre jours à Gatineau, du 10 au 13 juillet 2012. Approximativement 2000 cahiers ont été corrigés, dont environ 1000 cahiers en français et 1000 cahiers 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 14 correctrices et correcteurs, dont la moitié se trouvait à la table francophone et l’autre moitié à la table anglophone.

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

### **Capture des données**

Les élèves ont indiqué leurs réponses (pour les items à réponse choisie) sur un feuillet de réponses détachable et ont écrit leurs réponses (pour les items à réponse construite)

directement dans les cahiers d'évaluation du PPCE. Pour les items à réponse choisie, chaque réponse correcte recevait un point. Pour les items à réponse construite, qui pouvaient recevoir un plein crédit ou un crédit partiel, le score attribué se situait sur un barème de 0 à 2.

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

### **Analyse des données**

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

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

L'analyse des données s'est effectuée selon la théorie classique. Le comité consultatif technique du PPCE a utilisé les données ainsi obtenues pour sélectionner, d'un point de vue statistique, les meilleurs items pour l'évaluation proprement dite 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 bien répondu 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 pertinentes, comme le calcul des moyennes pour chaque item et la préparation des distributions des fréquences pour le pourcentage d'élèves ayant sélectionné chaque réponse pour les items à réponse choisie ou ayant obtenu chaque code pour les items à réponse construite. Ils ont également analysé le pourcentage de données manquantes et effectué une analyse du fonctionnement différentiel des items (FDI) selon le sexe et la langue.

## **Groupe de travail sur la sélection des items — Évaluation proprement dite**

Après la mise à l'essai des items, le groupe de travail sur la sélection des items, composé de représentantes et représentants de sept instances, s'est réuni pour examiner et sélectionner des scénarios et des items pour l'évaluation proprement dite. Tous les cahiers 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 les deux sexes. Ils ont aussi pris en compte les commentaires des correctrices et correcteurs qui avaient participé à la séance de correction des items de la mise à l'essai (à partir du questionnaire qui leur avait été donné à la fin de la séance de correction). Ce questionnaire contenait des remarques pertinentes sur l'instrument d'évaluation en général, de même que des commentaires sur chaque item concernant sa qualité.

Le groupe de travail a sélectionné les items pour les trois domaines : sciences, lecture et mathématiques. Un très petit nombre d'items en sciences des évaluations précédentes du PPCE ont été retenus, mais sans l'intention de les utiliser comme items d'ancrage. Tous les items en lecture et en mathématiques étaient des items d'ancrage et étaient donc identiques à ceux utilisés lors de l'évaluation où ils faisaient partie du domaine principalement : lecture pour le PPCE de 2007 et mathématiques pour le PPCE de 2010. 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 cahiers d'évaluation, afin de déterminer si les élèves réussissaient mieux pour un item dans une langue que dans l'autre et, le cas échéant, de discuter des raisons possibles de tels résultats.

Le groupe de travail a également examiné les réponses des questionnaires de l'élève, de l'école et du personnel enseignant au niveau du contenu et d'un point de vue statistique et psychométrique. Il a choisi les items qui, selon lui, pouvaient fournir le plus de données pertinentes lors de l'évaluation proprement dite, 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'évaluation proprement dite. Le *Cadre d'évaluation en sciences du PPCE* a été par la suite réexaminé afin de s'assurer qu'il correspondait bien aux items de l'évaluation. Très peu de changements ont été nécessaires.

## **Chapitre 6 – Évaluation proprement dite**

L'évaluation du PPCE a eu lieu entre le 29 avril et le 24 mai 2013, le domaine principal étant les sciences et les domaines secondaires étant la lecture et les mathématiques. Plus de 32 000 élèves, sélectionnés de façon aléatoire dans près de 1600 écoles dans 10 provinces du Canada, ont participé à l'évaluation en français et en anglais. Les principaux items ont permis d'évaluer les connaissances et les habiletés des élèves de 8<sup>e</sup> année/2<sup>e</sup> secondaire dans les trois matières.

### **Cahiers d'évaluation**

Chaque cahier d'évaluation contenait deux blocs de questions en sciences, un bloc d'items en lecture et un bloc d'items en mathématiques. Pour évaluer l'équivalence de chaque cahier, un sous-ensemble d'items de chaque domaine était répété dans des paires de cahiers. Lors de la mise en page des cahiers, 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 instances ont dû les examiner, de façon que leurs commentaires puissent être incorporés le cas échéant. Le matériel envoyé aux instances à des fins de réexamen comprenait toutes les versions des cahiers 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 instance pour le matériel d'évaluation.

### **Impression des cahiers d'évaluation**

Après que le CMEC a reçu l'approbation de toutes les instances, les cahiers d'évaluation ont été imprimés à des fins de vérification, dans l'optique de s'assurer que toutes les modifications de contenu apportées par les groupes de travail et le CMEC avaient été intégrées aux nouvelles versions des quatre cahiers d'évaluation. Une fois ce processus terminé, les cahiers d'évaluation ont ensuite été convertis au format PDF et imprimés. Un numéro d'identification unique muni d'un code à barres a été imprimé sur chaque cahier afin de l'attribuer à l'élève correspondant. Les cahiers d'évaluation ont ensuite été envoyés aux coordonnatrices et coordonnateurs des instances, qui devaient les distribuer aux écoles sélectionnées.

### **Vérification des documents**

Chaque coordonnatrice et coordonnateur des instances devait s'assurer qu'il disposait des documents destinés à chaque école. Tout colis manquant devait être immédiatement signalé au CMEC, afin de s'assurer que les documents seraient bien arrivés avant la date prévue de l'évaluation. Si les directions des écoles et les districts/conseils/commissions 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 coordinateur de leur instance directement.

## **Lettre envoyée aux parents/tuteurs des élèves**

Avant l'évaluation proprement dite, la coordonnatrice ou le coordonnateur de l'école devait informer les élèves sélectionnés ainsi que leurs parents/tuteurs ou tutrices. 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 coordinateur 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 l'instance.

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

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

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

## **Lieu de l'évaluation**

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

## **Déroulement de l'évaluation**

Au début de l'évaluation, la coordonnatrice ou le coordonnateur de l'école a distribué un exemplaire du cahier d'évaluation à chaque élève indiqué dans le formulaire de suivi des élèves. Il fallait distribuer à parts égales les quatre cahiers aux élèves de la classe. La coordonnatrice ou le coordonnateur devait aussi s'assurer d'avoir donné les instructions aux élèves avant le début de l'évaluation; ils 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 de temps supplémentaire 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 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 cahiers d'évaluation et d'établir le statut de l'élève, afin de déterminer si ce dernier avait participé à l'évaluation, en avait été exempté ou avait été absent.

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

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

## **Élèves ayant des besoins spéciaux**

Pour cette évaluation, les aménagements étaient définis 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'aménagements avaient dû être identifiés au préalable, lorsque l'école avait soumis sa liste d'élèves admissibles. Les coordonnatrices et coordonnateurs des écoles devaient aviser la coordonnatrice ou le coordonnateur de leur instance 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 cahiers d'évaluation à l'école. Il était important de prendre les dispositions nécessaires pour permettre aux élèves ayant des besoins spéciaux de participer autant que possible à l'évaluation sans risquer de compromettre l'intégrité de cette dernière.

Les aménagements étaient autorisés uniquement pour les élèves qui en bénéficient habituellement durant leurs travaux ordinaires en classe. La liste des aménagements autorisés comprenait les éléments suivants : braille, version en gros caractères, papier coloré et version audio. Ces aménagements étaient accordés 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 aménagements étaient également à la disposition de l'ensemble des élèves :

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

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

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

Le questionnaire de l'école était généralement rempli par la directrice ou le directeur de l'école. Les enseignantes et enseignants de sciences des classes sélectionnées devaient remplir le questionnaire du personnel enseignant. Les deux questionnaires étaient offerts en versions imprimée et en ligne. Dans certaines instances, il y avait quelques écoles qui étaient structurées de 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 sciences associés aux élèves sélectionnés étaient priés de remplir le questionnaire (à raison d'un questionnaire par enseignante ou enseignant). Chaque questionnaire comportait un numéro d'identification inscrit sur sa page couverture. Un numéro d'identification du questionnaire du personnel enseignant était attribué à chaque enseignante ou enseignant identifié et imprimé sur la page couverture du questionnaire. Les questionnaires devaient être distribués au moment de l'évaluation.

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

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

## Participation à l'évaluation et exemptions

On s'attendait à ce que les élèves de 8<sup>e</sup> année/2<sup>e</sup> secondaire aient les capacités nécessaires pour faire 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 proprement dite.

Il se pouvait, cependant, qu'on s'attende à ce 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 sciences, en lecture ou en mathématiques 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 et chaque territoire. Les trois codes d'exemption sont fournis ci-dessous.

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

1. 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'un tel aménagement en situation de test pendant le programme scolaire normal;
2. pause ou multiples pauses, à condition que l'élève soit supervisé pendant ces pauses;
3. autre local;
4. utilisation du braille, de gros caractères d'imprimerie ou de papier en couleur;
5. recours à un scribe (cette personne ayant l'obligation d'écrire exactement ce que dit l'élève, mot pour mot, sans modification);
6. lecture mot pour mot des instructions uniquement, dans tous les domaines;
7. lecture mot pour mot d'instructions ou de questions ponctuelles pour les sections de sciences et de mathématiques seulement (dans les cas où il faut lire en entier le test de sciences ou de mathématiques, une version audio du test sur disque compact peut être fournie).

- I =** Exemption pour cause d'incapacités intellectuelles ou de problèmes socioaffectifs. Un élève qui est considéré, selon l'opinion professionnelle de la direction de l'école ou d'autres membres qualifiés du personnel, comme ayant des incapacités intellectuelles ou des problèmes socioaffectifs ou dont les tests ont montré qu'il avait de telles incapacités ou problèmes devrait être exempté. Cette catégorie inclut les élèves qui sont incapables, sur le plan mental ou affectif, de suivre ne serait-ce que les instructions générales du test.
- N =** Exemption en raison de la langue (locutrices ou locuteurs non natifs). Cette exemption ne s'applique qu'aux élèves dont la langue première n'est ni le français ni l'anglais. Dans les évaluations de grande envergure, les écoles peuvent examiner le cas des élèves qui sont au Canada depuis moins de deux ans.

Le nombre et les pourcentages d'élèves exemptés se trouvent au tableau I-2 à l'Annexe I du rapport public (O'Grady & Houme, 2014)<sup>16</sup>.

### **Organisation d'une séance de rattrapage**

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

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

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

### **Renvoi du matériel d'évaluation**

Après l'évaluation des élèves, les coordonnateurs et coordonnatrices des écoles devaient remplir le « Rapport de la coordonnatrice ou du coordonnateur de l'école ». Ils devaient aussi remplir la « Liste des documents expédiés » et indiquer le nombre d'exemplaires de chaque document. Dès que possible après l'évaluation, ils devaient renvoyer à la coordonnatrice ou au coordonnateur de l'instance 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 cahiers d'évaluation et les feuilles de réponses, ainsi que les exemplaires et les photocopies des cahiers d'évaluation non utilisés.

---

<sup>16</sup> <http://cmec.ca/Publications/Lists/Publications/Attachments/337/PCAP-2013-Public-Report-FR.pdf>

## **Séance de correction**

La séance de correction des items de l'évaluation proprement dite a eu lieu à Moncton (Nouveau-Brunswick) du 8 au 19 juillet 2013. Tous les items de sciences, de lecture et de mathématiques ont été corrigés par des enseignantes et enseignants de ces domaines. La correction des items de sciences s'est étalée sur deux semaines, parce qu'il s'agissait du domaine principal et qu'il y avait un plus grand nombre d'items à corriger. La correction des items de mathématiques s'est déroulée uniquement pendant la première semaine, tandis que la correction des items de lecture a eu lieu pendant la deuxième semaine. En tout, 120 correctrices et correcteurs y ont participé, en englobant à la fois les francophones et les anglophones.

Au total, plus de 32 000 cahiers d'évaluation ont été corrigés, dont environ 8000 en français et 24 000 en anglais.

## **Mise en liasse des cahiers**

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

## **Feuilles de correction**

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

## **Guide de correction**

Avant la séance de correction, les correctrices et correcteurs se sont vu fournir un guide de correction contenant des informations sur la logistique de la séance de correction et décrivant les responsabilités du personnel du CMEC, des chefs de correction et des correctrices et correcteurs. Ce guide fournissait également des informations sur le traitement des cas spéciaux, comme les points de vue tendancieux de la part de la correctrice ou du correcteur ou les

suspensions de triche. Le « Formulaire de commentaires de la correctrice ou du correcteur », qui devait être rempli à la fin de la séance de correction, était également inclus dans ce guide.

### **Guide de codification**

Le guide de codification fournissait une introduction générale sur l'art d'attribuer des codes aux réponses et des détails sur les principes de la codification, 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 codification fournissait la catégorie de chaque question et une description de tous les codes possibles, ainsi qu'un éventail d'exemples de réponses susceptibles de recevoir un plein crédit ou un crédit partiel pour la question.

### **Chefs de correction**

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

### **Chefs de table**

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

### **Formation des correctrices et correcteurs**

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

façon cohérente, ils sont passés à un travail individuel de correction, jusqu'à ce que tous les cahiers d'évaluation soient corrigés. Une fois que la correction d'un scénario était achevée, les correctrices et correcteurs ont reçu une formation sur la correction des items du scénario suivant.

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

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

### **Fidélité des corrections**

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

### ***Examens de fidélité***

Dans le cadre d'une séance de correction, il est toujours important d'instaurer les procédures nécessaires pour s'assurer que les correctrices et correcteurs corrigent comme il se doit. Il s'agit d'une étape majeure, car ils doivent tous s'entendre sur les différents codes afin d'assurer la validité des résultats. Avant la séance de correction, le personnel du CMEC a choisi des items au hasard dans l'ensemble des cahiers d'évaluation reçus des instances pour faire des examens de fidélité. Ces items choisis dans un ou plusieurs scénarios ont ensuite été distribués aux chefs de correction pour qu'ils en fassent la codification. Par la suite, leurs réponses ont été renvoyées au personnel du CMEC, qui les a saisies dans un fichier Excel. Lors de la séance de correction, si les chefs de correction mettaient en évidence un problème spécifique sur des questions particulières, des examens de fidélité supplémentaires étaient élaborés pour régler le problème. Les examens de fidélité ont donc fait office à la fois de contrôle de la qualité et de formation supplémentaire pour les correctrices ou correcteurs. Les examens de fidélité ont été

exécutés pour l'ensemble des correctrices ou correcteurs francophones ou anglophones dans les trois domaines d'évaluation. Le but de ces examens était de surveiller la cohérence des corrections tout au long de la séance. Les examens de fidélité se sont produits environ deux fois par jour, selon la procédure suivante :

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

Les examens de fidélité ont donc permis de vérifier la concordance entre les résultats des experts et ceux des correctrices et correcteurs, c'est-à-dire qu'ils ont permis de vérifier si ces derniers avaient attribué les mêmes codes que les spécialistes pour les items d'un ou de plusieurs scénarios. Pour chaque examen de fidélité, il y avait un pourcentage du degré d'accord pour chaque correctrice et correcteur et pour chaque item. On s'attendait à ce que le degré d'accord entre les spécialistes (chefs de correction) et les correctrices et correcteurs soit d'environ 85 p. 100. Si le résultat global des examens de fidélité était faible pour telle ou telle question ou tel ou tel bloc de questions spécifique, alors le groupe suivait une nouvelle formation et les copies déjà corrigées étaient vérifiées à nouveau par les chefs de correction, ou bien la codification de la question recommençait. Si le résultat des examens de fidélité était faible pour des correctrices ou correcteurs ou pour des tables spécifiques, alors les chefs de table offraient une nouvelle formation à l'individu ou au groupe concerné avant de passer à la correction. Les items précédemment corrigés par ces correctrices ou correcteurs étaient vérifiés.

À la fin de la séance de correction, tous les pourcentages obtenus pour chaque examen de fidélité de chaque correctrice et correcteur ont été compilés, ce qui a constitué le degré d'accord total en pourcentage. Les résultats nous montrent que la majorité des correctrices et correcteurs ont obtenu un niveau d'accord en pourcentage avec les spécialistes qui était plus qu'acceptable. Les résultats aux examens de fidélité ont été satisfaisants pour les trois matières évaluées. Les quatre tableaux suivants présentent le niveau d'accord en pourcentage par langue et globalement pour chaque groupe de correction et chaque domaine.

**TABLEAU 6.1 – Résultats des examens de fidélité pour le groupe de correction en sciences n° 1**

	Bloc 1		Bloc 2		Bloc 3			Bloc 4				Niveau d'accord en %
	Test 1	Test 2	Test 1	Test 2	Test 1	Test 2	Test 3	Test 1	Test 2	Test 3	Test 4	
Anglais	90	90	88	99	83	93	-	85	94	98	97	92
Français	92	85	93	96	95	94	95	82	80	98	93	91
Global	91	88	91	98	89	94	95	84	87	98	95	91

**TABLEAU 6.2 – Résultats des examens de fidélité pour le groupe de correction en sciences n° 2**

	Bloc 1		Bloc 2				Bloc 3				Bloc 4				Niveau d'accord en %
	Test 1	Test 2	Test 1	Test 2	Test 3	Test 4	Test 1	Test 2	Test 3	Test 4	Test 1	Test 2	Test 3	Test 4	
Anglais	81	88	83	70	92	98	74	75	95	-	96	95	96	-	87
Français	99	98	91	94	91	98	82	87	96	88	86	93	88	96	92
Global	90	93	87	82	92	98	78	81	96	88	91	94	92	96	89

**TABLEAU 6.3 – Résultats des examens de fidélité pour le groupe de correction en lecture**

	Bloc 1				Bloc 2			Niveau d'accord en %
	Test 1	Test 2	Test 3	Test 4	Test 1	Test 2	Test 3	
Anglais	72	85	89	73	69	74	78	77
Français	83	84	79	95	81	89	88	86
Global	78	84	84	84	75	82	83	81

**TABLEAU 6.4 – Résultats des examens de fidélité pour le groupe de correction en mathématiques**

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

***Fiabilité entre membres du groupe de correction (double correction)***

La double correction est un indicateur de contrôle de la qualité pour la correction des cahiers d'évaluation en sciences, en lecture et en mathématiques; elle consiste à prendre environ 2400 cahiers (300 cahiers de chaque catégorie, en anglais et en français) et de les faire corriger une deuxième fois par une autre correctrice ou un autre correcteur. Comme le montre le tableau 6.5, le niveau global de concordance entre membres du groupe de correction est de 95 p. 100 à la fois en sciences et en mathématiques et de 80 p. 100 en lecture.

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

Cahier	Domaine et groupe de correction			
	Groupe de correction en sciences n° 1	Groupe de correction en sciences n° 2	Groupe de correction en lecture	Groupe de correction en mathématiques
<b>1</b>	99	98	92	99
<b>2</b>	95	95	81	97
<b>3</b>	95	94	71	93
<b>4</b>	90	92	77	90
<b>Moyenne</b>	95	95	80	95

### ***Fiabilité des tendances***

La fiabilité des tendances est un indicateur de contrôle de la qualité servant à estimer le niveau d'accord entre les correctrices et correcteurs en mathématiques et en lecture pour les items d'ancrage à l'évaluation du PPCE de 2013 et pour les mêmes items à l'évaluation du PPCE de 2010. Quatre items pour la lecture et neuf items pour les mathématiques étaient communs aux deux évaluations. Environ 2400 cahiers de 2010 (300 exemplaires de chaque cahier, en français et en anglais) ont été corrigés une deuxième fois par une autre correctrice ou un autre correcteur. Lorsque les items étaient présents dans plus d'un cahier, seul un cahier a été corrigé pour la fiabilité des tendances ou bien des nombres égaux de chaque cahier, jusqu'à un maximum de 300. Les cahiers ont été corrigés selon ce qu'avaient décidé les chefs de correction. La correction pour la fiabilité des tendances s'est faite en même temps que le travail de correction principal, mais au début du processus, afin de pouvoir utiliser les données pour faire concorder les scores entre les deux évaluations si cela s'avérait nécessaire.

### ***Correction multiple***

Lors de la correction du PPCE de 2013, les deux blocs d'items de sciences de chaque cahier ont été corrigés par différents groupes de correctrices et correcteurs. Afin de vérifier que la correction des deux groupes de correction était équivalente, l'un des cahiers a été corrigé par l'ensemble des correctrices et correcteurs.

### ***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 instances. Les correctrices et correcteurs ont fait des commentaires lors de la séance de correction. Les informations recueillies dans ces rapports sont résumées dans les sections qui suivent.

### ***Rapport de la coordonnatrice ou du coordonnateur de l'instance***

À la suite du déroulement de l'évaluation, les coordonnatrices et coordonnateurs des instances 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 instances 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 l'instance » incluait sept questions.

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

Le *Guide d'information à l'intention de l'école* décrivait les procédures administratives pour les tests du PPCE. Malheureusement, les responsables du test ne se sont pas tous familiarisés avec ce guide à l'avance. Le guide encourageait, par exemple, le personnel enseignant à accorder aux élèves de petites pauses si nécessaire tout au long de l'évaluation, mais les enseignantes et enseignants ne semblaient pas tous être au courant. 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 des synonymes. Malheureusement, il semble que ces choses n'aient pas été mises à la disposition des élèves dans toutes les écoles. Les écoles pouvaient choisir la langue de participation à l'évaluation pour les élèves d'immersion française, mais certaines écoles ont indiqué que les élèves d'immersion ayant participé à l'évaluation en anglais n'étaient pas bien familiarisés avec le vocabulaire du test, parce que l'enseignement en salle de classe s'était déroulé en français.

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

coordonnateurs des instances sont parvenus à régler certains de ces problèmes quand ils se sont présentés, parce qu'ils avaient des cahiers supplémentaires dans les deux langues.

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

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

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

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

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

### ***Rapport de la coordonnatrice ou du coordonnateur de l'école***

Après l'évaluation, les coordonnatrices et coordonnateurs des écoles devaient remplir un rapport pour indiquer si la séance d'évaluation s'était bien déroulée. Le « Rapport de la coordonnatrice ou du coordonnateur de l'école » incluait 13 questions. Les commentaires des coordonnatrices et coordonnateurs aideront à mieux planifier le déroulement des évaluations à l'avenir et ont aidé à recueillir des informations sur la manière dont l'évaluation du PPCE de 2013 s'est déroulée dans les écoles.

Tout d'abord, les coordonnatrices et coordonnateurs des écoles devaient décrire les mesures prises pour inciter les élèves à prendre au sérieux leur participation à l'évaluation. Les mesures utilisées pour encourager les élèves à faire de leur mieux étaient semblables à celles décrites précédemment par les coordonnatrices et coordonnateurs des instances. Plusieurs coordonnatrices et coordonnateurs des écoles ont mentionné qu'ils avaient passé du temps avec les élèves sélectionnés pour les sensibiliser à l'évaluation et leur expliquer l'importance de faire de leur mieux au test et de répondre aux questions de façon sérieuse. Certains d'entre eux ont même rencontré les élèves quelques jours avant l'évaluation et passé en revue les exemples de questions avec eux. D'autres leur ont expliqué que les données obtenues permettraient de comparer les résultats des différentes instances et qu'il était important de bien représenter son école et sa province ou son territoire. Certains autres ont envoyé des informations sur l'évaluation du PPCE aux parents et tuteurs ou tuteurs des élèves sélectionnés pour l'évaluation, pour qu'ils sachent quel était le but de l'évaluation et qu'ils puissent eux-mêmes motiver leur enfant. Certains coordonnateurs et coordonnatrices des écoles ont mentionné que certains élèves avaient été récompensés à la fin de l'évaluation par une collation gratuite, comme une pizza, ce qui a semblé être une source de motivation les encourageant à faire de leur mieux au test.

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

Les coordonnatrices et coordonnateurs des écoles devaient aussi indiquer s'il y avait eu des problèmes pendant le déroulement de l'évaluation. La plupart d'entre eux ont mentionné que la séance d'évaluation s'était bien déroulée, même si, pour certains, divers problèmes sont survenus lors de l'évaluation. Avant l'évaluation, les coordonnatrices et coordonnateurs des écoles ont reçu la liste des élèves dans le formulaire de suivi des élèves, avec un numéro d'identification pour chacun. Certains surveillants et surveillantes d'examen n'ont pas lu attentivement les consignes et ont distribué les cahiers d'évaluation de façon aléatoire et ne se sont rendu compte que plus tard que les cahiers portaient des numéros les assignant à des élèves spécifiques. Il y a donc eu de la confusion pour savoir à qui appartenait chaque cahier. Ils ont aussi remarqué que certains élèves étaient fatigués et n'étaient pas motivés pour faire le test, car l'évaluation avait lieu à la fin de l'année (la participation aux tests ayant eu lieu, dans

plus de 80 p. 100 des cas, au mois de mai) et les élèves avaient plusieurs examens de fin d'année. Par conséquent, certains coordonnateurs et coordonnatrices des écoles ont exprimé leur crainte que certains élèves n'aient pas pris le test au sérieux. D'autres coordonnatrices et coordonnateurs des écoles ne savaient pas que les élèves pouvaient avoir accès à des ressources matérielles (p. ex., dictionnaire unilingue ou bilingue, articles à manipuler, dictionnaire des synonymes ou calculatrice) et ont recommandé de mieux le préciser avant le déroulement du test. Certaines écoles ont indiqué que la voix dans l'enregistrement sur disque compact en anglais avait un accent difficile à comprendre pour certains élèves.

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

- Certaines écoles ont choisi de ne pas aviser les élèves ni les parents et tutrices ou tuteurs du test avant la date d'évaluation.
- Dans un cas de figure, une personne chargée de surveiller le test n'avait pas d'adresse à laquelle renvoyer les copies et a indiqué qu'elle s'était débarrassée du test.
- Par inadvertance, certaines écoles n'ont pas tenu compte de la consigne relative aux numéros de code d'identification des cahiers d'évaluation (soit la correspondance entre les codes d'identification spécifiques et les noms des élèves). Une fois les problèmes cernés, les erreurs ont néanmoins été corrigées.
- Certaines personnes chargées de la surveillance ont perdu les formulaires de suivi des élèves et distribué les cahiers d'évaluation au hasard. Il y a du coup eu de la confusion pour savoir à qui appartenait chaque cahier.
- Certaines écoles n'ont pas rempli correctement les formulaires de suivi des élèves, ce qui a créé des problèmes par la suite.
- Dans quelques écoles, du temps supplémentaire a été accordé au-delà de ce qui était permis dans les lignes directrices de l'évaluation.
- Du fait que plusieurs écoles ont estimé que la durée du test était trop longue par rapport au niveau de concentration de certains élèves ou en raison d'événements spéciaux qui avaient lieu à l'école, l'évaluation a été divisée en deux ou trois séances, au lieu de se dérouler en une seule séance de 90 minutes ou plus avec de petites pauses.
- Dans quelques écoles, la classe sélectionnée n'était pas une classe de sciences. Les questions relatives à l'enseignement et à l'apprentissage des sciences dans le questionnaire du personnel enseignant ont posé des problèmes à l'enseignante ou enseignant concerné.

Les coordonnatrices et coordonnateurs des écoles devaient commenter les scénarios d'introduction (p. ex., pertinence, niveau de difficulté, niveau d'intérêt) et les questions de l'évaluation (p. ex., formulation maladroite, plus d'une réponse correcte ou aucune réponse adéquate, question non adaptée à l'âge des élèves). La plupart des coordonnatrices et coordonnateurs des écoles ont précisé que le niveau des questions de l'évaluation était approprié et que les élèves se sont sentis engagés. Toutefois, plusieurs ont estimé que les

questionnaires étaient trop longs et parfois répétitifs. Par ailleurs, un certain nombre de commentaires ont été faits à propos de la formulation des questions de l'évaluation.

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

### ***Formulaire de commentaires de la correctrice ou du correcteur***

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

Au total, 58 p. 100 des personnes interrogées ont corrigé les items de sciences, qui constituaient le domaine principal de l'évaluation; 21 p. 100 ont corrigé les items en lecture; et 21 p. 100 ont corrigé les items en mathématiques. Ces deux dernières matières étaient les domaines secondaires lors de l'évaluation du PPCE de 2013. En outre, environ 10 p. 100 des correctrices et correcteurs ont corrigé des copies dans deux domaines.

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

On demandait aux correctrices et correcteurs d'examiner les questions d'évaluation afin de faire part de leurs observations, lors de la séance de correction, sur les forces et les faiblesses des questions en sciences. On leur demandait aussi de dresser une liste d'idées fausses de la part des élèves dans leurs réponses. Les informations recueillies ont servi à préparer un numéro de *L'évaluation... ça compte!*, dans lequel 14 items de l'évaluation du PPCE de 2013 en sciences ont été publiés, avec des commentaires sur les travaux des élèves et des exemples de réponses<sup>17</sup>.

---

<sup>17</sup> Numéro disponible à l'adresse [http://cmec.ca/Publications/Lists/Publications/Attachments/339/AMatters\\_No8\\_PCAPItems\\_FR.pdf](http://cmec.ca/Publications/Lists/Publications/Attachments/339/AMatters_No8_PCAPItems_FR.pdf).

## **Chapitre 7 – Établir une norme de rendement**

Chaque fois que le contenu ou les types d'items utilisés dans les tests subissent des modifications significatives, il convient d'effectuer un travail de normalisation. 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 d'ajustements 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 2013, les sciences ont été le domaine principal du PPCE pour la première fois et des changements significatifs ont été apportés au cadre d'évaluation, de sorte qu'il a été nécessaire d'établir des normes de rendement.

### **Séances de normalisation**

Les séances d'établissement des normes ont eu lieu du 25 au 28 novembre 2013 à Toronto. Les réunions ont été divisées en trois séances : une séance de formation des chefs d'une journée, le 25 novembre; deux journées de séances de normalisation, les 26 et 27 novembre; et une séance de rédaction d'une journée, pour réviser les descripteurs des niveaux de compétence, le 26 novembre.

La normalisation avait pour but de définir les niveaux de rendement pour l'évaluation en sciences du PPCE de 2013 pour les élèves de 8<sup>e</sup> année/2<sup>e</sup> secondaire. 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. Trois points de coupure ont été définis pour distinguer quatre niveaux de rendement. Le niveau 2 a été désigné comme le niveau de rendement acceptable pour les élèves de 8<sup>e</sup> année/2<sup>e</sup> secondaire.

Les participants ont fait les tests, les ont corrigés, ont examiné les descripteurs des niveaux de rendement (DNR) et se sont ensuite livrés à trois cycles d'examen des tests à l'aide de la méthode du signet (Cizek et Bunch, 2007). À l'issue des quatre journées, les points de coupure recommandés par les spécialistes ont été envoyés aux coordonnatrices et coordonnateurs des instances 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 instances soient impliquées dans le processus et qu'elles aient l'occasion de participer à la définition des points de coupure. Chaque instance a été invitée à désigner deux représentantes ou représentants ayant une certaine expertise en mesure et en évaluation et en ce qui concerne le contenu du programme de sciences pour la 8<sup>e</sup> année/2<sup>e</sup> secondaire. Le comité de normalisation se composait de 25 spécialistes. Le CMEC a recruté les membres du comité et les spécialistes responsables de la normalisation en s'appuyant sur les noms fournis par les coordonnatrices et coordonnateurs des instances. L'un des critères les plus importants pour un tel groupe de travail est qu'il faut qu'il présente 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. Le groupe de spécialistes se composait également d'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 de normalisation. Ces DNR décrivent ce que les élèves devraient savoir et être capables de faire à chacun des quatre niveaux de rendement pour la 8<sup>e</sup> année/2<sup>e</sup> secondaire. Les DNR sont cruciaux dans le processus de normalisation, parce qu'ils guident les spécialistes en les aidant à conceptualiser les différences de niveau de rendement entre élèves.

La rédaction des versions préliminaires des DNR a été influencée par les travaux de recherche sur les tests internationaux (TEIMS, PISA, etc.). 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. Les DNR ont été énoncés dans des termes correspondant au cadre d'évaluation en sciences du PPCE et, à cette première étape, les spécialistes ont fait des suggestions de révisions à apporter, en se consultant les uns les autres et sur les conseils de la personne responsable du dossier au CMEC. Lors de la réunion, la personne responsable a pris en note les révisions et les suggestions et les a projetées à l'écran, afin que tout le monde puisse bien les voir. Elle a ensuite intégré les formules suggérées pour produire des DNR cohérents, sur lesquels les spécialistes se sont mis d'accord. À l'issue de ce processus, les DNR étaient prêts à être utilisés lors de la réunion de normalisation.

### **Sécurité des documents**

Comme l'établissement des normes utilise des documents authentiques, il était crucial d'assurer un bon niveau de sécurité. À son arrivée à l'atelier, chaque spécialiste a reçu un code d'identification qui lui était propre. Tous les documents sécurisés contenaient les mêmes codes, pour que, lors de la distribution, l'exemplaire du document corresponde au numéro du spécialiste auquel il était distribué. Les spécialistes ont été informés du fait 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 plusieurs 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 de 2013 combinait des items à réponse choisie (c'est-à-dire 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 sciences auxquels les élèves pouvaient répondre par une réponse brève, à laquelle était attribué le code 1 ou 0, selon qu'elle était correcte ou incorrecte. Les items à RCSL étaient des items comportant un ou deux points et exigeant des élèves une réponse écrite plus longue.

**TABEAU 7.1 – Composition des cahiers d'évaluation du PPCE pour les items de sciences**

	Réponse choisie	Réponse construite courte (code 0 ou 1)	Réponse construite longue (code 0, 1 ou 2)
Cahier 1	17	5	2
Cahier 2	18	4	2
Cahier 3	17	6	1
Cahier 4	17	5	2

Avec la méthode du signet, les spécialistes ont examiné les items du test dans un cahier des items ordonnés (CIO), dans lequel tous les items des quatre cahiers 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 CIO, 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. Chaque page contenait des informations essentielles sur l'item, comme sa position dans le CIO, sa position dans son cahier d'origine et le niveau de rendement (thêta) exigé pour que la probabilité que l'élève donne la réponse correcte ou obtienne le point soit de deux sur trois ou plus.

Le CMEC a utilisé le modèle de Rasch pour le calibrage des items et la construction du test. Ce modèle permet le calibrage de tous les items et des élèves selon une échelle commune. Ce calibrage commun permet de calculer la probabilité d'une réponse correcte à un item donné pour un élève donné à partir d'informations sur le niveau de rendement de l'élève (thêta) et le niveau de difficulté de l'item (valeur  $p$ ).

## Procédure de normalisation

Au total, 25 participantes et participants de l'ensemble des instances et deux membres du personnel du CMEC ont pris part à la séance de définition des points de coupure. Les participantes et participants ont été affectés à l'une des cinq tables prévues, soit deux tables francophones, deux tables anglophones et une table bilingue. Chaque table avait un chef et cinq participantes ou participants. Le processus de définition des points de coupure a pris deux journées, avec une troisième journée réservée au peaufinage des descripteurs des niveaux de rendement.

Le groupe a suivi une présentation sur le PPCE, les procédures administratives, les caractéristiques des items et le cadre d'évaluation (ces informations étant tout particulièrement pertinentes pour les participantes et participants prenant part pour la première fois à un projet lié aux évaluations pancanadiennes du CMEC), ainsi que sur les points de coupure, la méthode du signet, les niveaux de rendement, l'emploi du temps pour la séance et les documents. Les participantes et participants n'avaient, pour la plupart d'entre eux, 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 corriger leurs réponses, de façon à se faire une meilleure idée des descripteurs des niveaux de rendement.

La première ronde de placement des signets a eu lieu avant la fin de la journée : les participantes et participants ont examiné chaque item dans le CIO, 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 cahier. Après cette discussion, chaque participante ou participant devait choisir un point de coupure – c'est-à-dire placer un signet dans le CIO à l'endroit précis où se trouvait la dernière question à laquelle un élève affichant un rendement au niveau précisé aurait deux chances sur trois de répondre correctement. Le CMEC a compilé les réponses de l'ensemble des participantes et participants en notant dans un fichier Excel les numéros des items pour les trois points de coupure. C'est la médiane de l'ensemble des réponses qui a défini le 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. Le personnel du CMEC a affiché les résultats (c'est-à-dire les numéros de tous les items marqués d'un signet et la médiane utilisée pour établir chacun des trois points de coupure). Il y avait de grandes variations évidentes entre les réponses des participantes et

participants, certains situant le premier point de coupure au tout début du CIO et d'autres le situant bien plus loin dans le cahier. Ces variations ont débouché sur d'importantes discussions pertinentes, les membres du groupe expliquant les uns aux autres pourquoi ils avaient placé le signet sur tel ou tel item particulier. Plusieurs des participantes et participants ont indiqué qu'ils avaient eu de la difficulté à placer le premier signet, parce qu'ils trouvaient que certains items étaient plus faciles pour les élèves alors que les données indiquaient le contraire. Ces items ont par conséquent été situés à un rang plus avancé dans le cahier. Pour le deuxième point de coupure, il y avait une différence encore plus grande entre la proposition la plus basse et la proposition la plus haute. Cependant, le groupe a eu moins de difficultés à placer le troisième point de coupure entre les niveaux 3 et 4. Un nombre nettement plus élevé de participantes et participants a choisi les mêmes items et il y a eu moins de variations. La première ronde a été un bon exercice pour les participantes et participants et leur a donné l'occasion d'échanger des commentaires et des opinions.

La deuxième ronde a été semblable à la première, avec les participantes et participants plaçant des signets dans le CIO pour déterminer les trois points de coupure et fournissant la justification de leur choix. Le personnel du CMEC a compilé les résultats. 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 trois points de coupure que lors de la première ronde. De même, lors de la deuxième ronde, il a semblé y avoir moins de variations dans le placement du deuxième point de coupure. Pour la deuxième ronde (contrairement à la première), les participantes et participants se sont vu présenter des données d'impact sur l'incidence de leurs choix, c'est-à-dire le pourcentage d'élèves se situant aux niveaux 1, 2, 3 et 4. Selon les réponses fournies par le groupe lors de la deuxième ronde, 2 p. 100 environ des élèves se situaient au niveau 1, 33 p. 100 au niveau 2, 53 p. 100 au niveau 3 et 11 p. 100 au niveau 4. 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. Les tableaux de la fréquence cumulée de la théorie des réponses aux items (TRI) pour la statistique thêta de chaque cahier 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-delà<sup>18</sup>. La valeur thêta moyenne de tous les cahiers est la statistique utilisée pour mesurer l'incidence. Des diagrammes à barres utilisant un code de couleurs ont été préparés pour illustrer la répartition des résultats pour chacun des points de coupure et ce, pour la deuxième et la troisième ronde de placement des signets. 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 CIO pour la dernière fois, en gardant ou en modifiant leurs choix précédents.

---

<sup>18</sup> Comme nous l'avons décrit plus haut, la statistique thêta a été ajustée pour la probabilité de réponse de deux sur trois.

D'après les réponses du groupe lors de cette troisième ronde, les pourcentages d'élèves se situant à chaque niveau de rendement se présentent sur le tableau 7.2.

**TABLEAU 7.2 – Répartition des élèves par niveau de rendement en sciences**

Niveau de rendement			
Niveau 1	Niveau 2	Niveau 3	Niveau 4
8 %	44 %	39 %	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 de normalisation 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 d'entre eux, qu'ils étaient satisfaits de l'organisation de la séance et des chefs et responsables de l'animation et ils ont fait des commentaires favorables sur l'instrument d'évaluation. La plupart ont déclaré que les textes et les questions étaient appropriés et que l'évaluation en sciences était équitable pour des élèves de 8<sup>e</sup> année/2<sup>e</sup> secondaire.

### **Descripteurs des niveaux de rendement**

Après le processus de normalisation, un sous-groupe du groupe de spécialistes s'est chargé de réviser 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 élaboré, à partir de ces items, une description des connaissances et des habiletés caractérisant le rendement des élèves à chacun des quatre niveaux<sup>19</sup>. Le niveau 2 est considéré comme le niveau acceptable ou « seuil de compétence », c'est-à-dire le niveau à partir duquel les élèves commencent à montrer qu'ils possèdent les compétences qui leur permettent de faire face aux situations de la vie courante en rapport avec les sciences. 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 sciences globalement et à chaque compétence et sous-domaine de l'évaluation.

---

<sup>19</sup> Ces descriptions figurent dans le rapport public sur le PPCE de 2013 (O'Grady & Houme, 2014), à l'adresse <http://cmec.ca/Publications/Lists/Publications/Attachments/337/PCAP-2013-Public-Report-FR.pdf>.

## **Chapitre 8 – Traitement des données du PPCE**

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

### **Collecte des données**

Les cahiers d'évaluation et les questionnaires ont été distribués lors de l'évaluation proprement dite. Au Canada, ce sont 32 604 élèves de 8<sup>e</sup> année/2<sup>e</sup> secondaire qui ont fait le test et répondu au questionnaire de l'élève. Au total, 1594 enseignantes et enseignants de sciences responsables des élèves participants ont répondu au questionnaire du personnel enseignant et 1917 directrices et directeurs ont répondu au questionnaire de l'école. La collecte de données pour ces documents s'est déroulée sur une période de plusieurs semaines.

### **Saisie des données**

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

À l'issue du déroulement de l'évaluation proprement dite, les instances ont toutes envoyé les cahiers d'évaluation, les feuilles de réponses et les questionnaires à Toronto pour la saisie des données. Le questionnaire de l'élève, le questionnaire du personnel enseignant et le questionnaire de l'école contenaient des items à réponse choisie et n'avaient pas à être codifiés par des spécialistes. Ils ont donc été envoyés à une entreprise externe pour la saisie des données. Les cahiers d'évaluation ont ensuite été expédiés à Moncton, au Nouveau-Brunswick, pour la séance de correction, lors de laquelle les correctrices et correcteurs ont corrigé tous les items à réponse construite dans plus de 32 000 cahiers. Les correctrices et correcteurs devaient attribuer un code à chaque item en noircissant les cercles appropriés sur une feuille de correction.

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

Pour les données sur le rendement, les fichiers contenant des données illisibles ou les items avec de multiples réponses ont été analysés par un logiciel de reconnaissance optique des marques. Si, par exemple, les cercles d'une feuille de réponses donnée n'étaient pas suffisamment noircis, le programme indiquait qu'il s'agissait d'un fichier problématique.

L'agente responsable des données a vérifié ces fichiers électroniques un par un et a fait la saisie des données manuellement.

### **Contrôle de la qualité des données saisies**

Pour les données des questionnaires, l'entreprise chargée de la saisie des données a programmé des règles spécifiques pour chaque section et chaque réponse aux questions dans les trois questionnaires. Tous les 25 questionnaires, une vérification de contrôle de la qualité a été effectuée, lors de laquelle la superviseure ou le superviseur a vérifié les échantillons en saisissant à nouveau le questionnaire. Les différences, s'il y en avait, ont été signalées à l'opératrice ou opérateur avant de procéder à une nouvelle saisie du questionnaire et à la saisie de nouveaux lots de questionnaires. Une fois que tous les lots de questionnaires ont été saisis et traités, des programmes ont validé le résultat produit en utilisant les règles établies avant la saisie des données.

### **Nettoyage des données**

Lorsque les données ont été présentées par les instances, 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 les variables supprimées, les variables ajoutées et les variables pour lesquelles les règles avaient été modifiées. Toutes les déviations ont été vérifiées et confirmées auprès des instances. Les fichiers de données ont ensuite été envoyés au centre de traitement des données du CMEC, pour faire l'objet de procédures de nettoyage et de recodification spécifiques.

### **Recodification générale**

Après que le centre de données du CMEC a mené l'enquête sur toutes les déviations 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 instance, tout au plus) :

- Les incohérences non résolues concernant l'identification des élèves et des écoles ont conduit à la suppression de l'enregistrement dans la base de données.
- Les enregistrements pour les élèves qui ne contenaient pas à la fois les données sur le rendement de l'élève et les données du questionnaire ont été supprimés.
- Lorsqu'une variable pour un item avait été supprimée dans le questionnaire en raison d'une erreur de traduction, elle a été supprimée.
- Les enregistrements doubles de données ont été repérés et seul un des enregistrements a été gardé, en se fondant sur des règles établies par le comité consultatif technique. Si, par exemple, deux fichiers portaient le même code d'identification, le fichier contenant moins d'informations a été supprimé.

### **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 cahiers, etc.). Le CMEC a analysé les données d'échantillonnage et de suivi, les a vérifiées et, si nécessaire, a procédé à une recodification supplémentaire. Si, par exemple, une instance 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 recodifiés 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 par un spécialiste externe. Les bases de données ainsi obtenues ont été envoyées au CMEC, avec certaines analyses préliminaires. 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 fournies. Ceci a permis de produire des informations sur les items du test qui semblaient avoir produit des résultats inacceptables et sur toutes les données ambiguës qui subsistaient dans les questionnaires. Avec ces informations, on a corrigé la clé et procédé à une recodification supplémentaire des données ambiguës. Par exemple, si l'ambiguïté était liée à des erreurs d'impression ou de traduction, alors un code « sans objet » a été appliqué à l'item.

Le résultat de la recodification (exigée 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 fichiers de données ont ensuite été pondérés en fonction de la taille des populations décrite dans le cadre d'échantillonnage du PPCE.

## Chapitre 9 – Analyse des données sur le rendement des élèves

Ce chapitre décrit l'analyse des données sur le rendement des élèves du PPCE de 2013. Il décrit et fournit un calendrier détaillé de l'exécution et de la coordination des différentes tâches. Le plan d'analyse a compris les étapes suivantes :

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

### **Analyse préliminaire**

L'analyse préliminaire est venue s'inscrire en prolongement du processus de nettoyage des données. Elle a compris trois étapes : (1) filtrage des données ; (2) recodification des items ; (3) gestion des données manquantes. Ces étapes ont été effectuées pour chaque cahier, en décomposant les informations selon l'instance 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 l'instance. Ces tableaux ont été utilisés :

- pour vérifier la présence d'anomalies dans les données (données aberrantes, clés incorrectes, etc.) ;
- pour examiner (examen de premier niveau) la répartition des réponses ;
- pour déterminer (et, au bout du compte, évaluer) le taux de données manquantes par item et par cahier<sup>20</sup>.

### ***Recodification des items***

Les données de l'évaluation du PPCE de 2013 comprenaient à la fois des réponses valables et des réponses non valables pour les items du test. Pour un item à choix multiples (CM), la réponse était valable si l'élève n'avait choisi qu'une option pour la réponse, que son choix soit correct ou non. La réponse était considérée comme non valable si l'élève avait choisi plus d'une réponse. Pour les questions ouvertes, la réponse construite (RCS) était traitée comme étant

---

<sup>20</sup> Les types de données manquantes et leur traitement sont décrits plus loin.

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.

Les items à CM des versions française et anglaise ont été recodifiés séparément avant la fusion des deux ensembles de données. Cela s'est avéré nécessaire parce que les clés pour certains des items à CM en lecture, qui étaient des items d'ancrage repris des évaluations précédentes, n'étaient pas les mêmes dans les deux langues parce que les leurres étaient présentés hiérarchiquement sous la forme d'une pyramide. Si l'on n'avait pas recodifié ces items, cela aurait débouché sur des problèmes lors du processus de calibrage (p. ex., avec l'impossibilité d'atteindre la convergence).

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

### ***Données manquantes***

Comme pour les autres évaluations à grande échelle, trois types de données manquaient dans l'évaluation du PPCE de 2013<sup>21</sup> :

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

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

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

### ***Items non utilisés dans l'évaluation***

Les items *non utilisés dans l'évaluation* découlent de la conception de l'évaluation du PPCE, qui s'appuie sur la technique d'échantillonnage à matrices multiples. Cette technique divise les items de l'évaluation en sections ou cahiers, avec certains items communs à certaines ou à l'ensemble des sections. Chaque section est ensuite affectée à un sous-groupe distinct dans

---

<sup>21</sup> 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.

l'échantillon principal. Au PPCE, les questions ont été divisées en quatre cahiers, avec certains blocs d'items communs à des paires de cahiers. Comme chaque élève ne devait faire que certains des items du test, il n'y avait pas de réponse pour les items affectés aux trois autres cahiers et les réponses étaient donc manquantes en raison de la conception même de l'évaluation. Du coup, 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, de sorte qu'il était impossible pour l'élève d'y répondre. Par exemple, il pouvait y avoir eu une erreur de photocopie ou d'impression rendant la question illisible. Le code « sans objet » n'a été utilisé que dans un petit nombre de cas et les items ont été traités comme des valeurs manquantes.

### *Items non atteints*

Les items non atteints correspondent aux questions auxquelles l'élève n'a pas répondu et qui étaient regroupées vers la fin de l'évaluation. Cette situation se produit dans le vecteur de réponses de l'élève parce qu'il n'a pas eu le temps d'y répondre. Dans les évaluations internationales, un item est considéré comme non atteint quand l'élève n'y a pas répondu et n'a pas non plus répondu à l'item qui le précède immédiatement. Il faut en outre que l'élève n'ait tenté de répondre à aucun autre item dans le reste du cahier<sup>22</sup>. 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 de 2013 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 indique qu'il convient de ne pas ignorer les items omis (cité dans de Ayala, 2009, p. 150). Pour lui, lorsqu'on ignore les items

---

<sup>22</sup> 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 cahier » (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 codifiée comme étant manquante » (OCDE, 2012, p. 236 dans le rapport en anglais).

omis, on risque d'obtenir une estimation de niveau de compétence élevée si l'élève ne répond qu'aux questions pour lesquelles il a confiance en sa capacité de trouver la bonne réponse. 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. Ce type de réponse a reçu le code 8.

### **Analyse des items**

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

### ***Analyse de la théorie classique des items***

L'objectif de l'analyse classique était de produire des statistiques pour un deuxième examen des items du PPCE de 2013. Pour le domaine principal, c'est-à-dire les sciences, le premier examen s'est appuyé sur les données de la mise à l'essai. Les domaines secondaires consistaient en des items d'ancrage tirés des évaluations précédentes. Les items d'ancrage ont été utilisés pour les domaines secondaires afin d'évaluer les changements pour ces items au fil du temps (ou d'une cohorte à l'autre) en fonction de l'estimation de leur niveau de difficulté. Les items d'évaluation en lecture ont été utilisés à la fois en 2007 et en 2010 et les items d'évaluation en mathématiques ont été utilisés en 2010. Ces items ont fait l'objet d'une mise à l'essai lorsque la lecture et les mathématiques étaient le domaine principal de l'évaluation. Les statistiques ont été examinées en guise de préparation à la sélection des items à inclure dans le PPCE de 2013.

L'analyse de la théorie classique des items 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 multiples, 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 correction pour les items à RCS).

Ces statistiques ont été calculées pour chaque cahier : quatre cahiers pour la version anglaise du test et quatre cahiers pour la version française. Il y a donc eu huit tests pour le domaine des sciences. Il y a également eu huit tests, quoique de moindre ampleur, pour chaque domaine secondaire. Pour ces domaines secondaires, le placement des items dans l'évaluation du

PPCE 2013 était conforme à leur placement dans l'évaluation initiale. L'effet de leur position a néanmoins lui aussi été évalué.

#### *Difficulté des items*

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

#### *Discrimination des items*

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

#### *Statistiques spécifiques pour les items à CM*

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

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

#### *Statistiques spécifiques pour les items à RCS*

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

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

#### *Examen pour mettre en évidence les données manquantes*

Pour chaque item, les aspects suivants ont été examinés :

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

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

### *Fiabilité de l'évaluation du PPCE de 2013*

Pour chaque domaine et sous-domaine, l'indice de cohérence interne (l'alpha de Cronbach) a été calculé sur l'ensemble des cahiers 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 instance.

### *Items problématiques*

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

- corrélation bisérielle de point inférieure à 0,20 ;
- valeur  $p$  inférieure à 0,20 ;
- valeur  $p$  égale ou supérieure à 0,85 ;
- items plus faciles ou plus difficiles pour une province par rapport à la moyenne nationale<sup>23</sup> ;
- corrélation bisérielle de point positive pour plus d'un leurre dans les items à CM ou corrélations bisérielles de point non ordonnées sur les différents niveaux pour les items à RCS ;
- moins de 5 p. 100 des élèves sélectionnant un des leures dans les items à CM ;
- moins de 10 p. 100 des élèves obtenant la valeur du score pour un item à RCS ;
- accord entre membres du personnel de correction de moins de 70 p. 100 pour la valeur du score pour un item à RCS.

### **Analyse de la TRI**

L'analyse de la TRI a fait intervenir les aspects suivants : (1) évaluation de la dimensionnalité du PPCE de 2013 ; (2) estimation des paramètres des items ; et (3) évaluation du caractère convenable des modèles de la TRI. L'évaluation des modèles de la TRI a fait intervenir la dépendance locale des items (DLI), la concordance entre la fonction mathématique du modèle et les données et l'invariance du PPCE de 2013. Le processus s'est achevé par une évaluation des données sur la validité du PPCE. L'évaluation de la validité s'est faite grâce au fonctionnement différentiel des items (FDI) et les données sur la validité ont été décrites dans le cadre de l'analyse du FDI.

### **Évaluation de la dimensionnalité du PPCE de 2013**

La dimensionnalité du PPCE de 2013 a été évaluée par analyse factorielle des items (AFI). L'AFI désigne la classe d'approches non linéaires pour déterminer la structure factorielle des données catégoriques (Cai, 2010). Ces approches sont plus appropriées que l'analyse factorielle classique, qui se fonde sur une matrice de corrélation linéaire entre les variables observées. Cette approche linéaire débouche sur l'extraction d'éventuels facteurs factices lorsqu'on a

---

<sup>23</sup> Ceci suppose que le modèle de Rasch est ajusté aux données afin de repérer les items et que l'analyse d'interaction selon la province est exécutée.

affaire à des variables dichotomiques (ou polytomiques) (de Ayala, 2009 ; Laveault et Grégoire, 2002). Les approches non linéaires correspondent mieux, par conséquent, à ces types de données que les approches linéaires (McDonald, 1967).

Deux programmes de statistique, EQSIRT et IRTPRO, ont 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.

En sus de la méthode MVIC, la méthode *Normal Ogive Harmonic Analysis Robust Method – NOHARM* (analyse harmonique pour le modèle ogive normal) de Fraser et McDonald (2003) a également été utilisée. Cette méthode a produit de bons résultats dans les études de recouvrement de la dimensionnalité (de Ayala, 2009 ; voir aussi De Champlain et Gessaroli, 1998 ; Finch et Habing, 2005 ; Knol et Berger, 1991). Les trois différents types de logiciels s’entendaient sur une dimension dominante sous-tendant le domaine principal du PPCE de 2013<sup>24</sup>.

### **Calibrage des items**

Des items de paires de cahiers ont fait l’objet d’un calibrage simultané afin d’établir des liens entre les cahiers et d’utiliser une échelle commune à la fois pour les items et pour les élèves. Cette procédure rend possible l’estimation des scores thêta sans que cela dépende de l’ensemble d’items auquel les élèves ont répondu.

Trois calibrages séparés des items pour les trois domaines ont été exécutés, avec estimation séparée des paramètres des items en lecture, en mathématiques et en sciences. En outre, le processus de calibrage pour la lecture et les mathématiques a utilisé les données de 2010 (évaluation dans laquelle les mathématiques étaient le domaine principal) et de 2013 simultanément<sup>25</sup>.

Trois modèles de la TRI ont été ajustés aux données simultanément. Pour les items à CM, la modélisation a ajusté le modèle à deux paramètres (2-PLM). Elle a ensuite été comparée au modèle concurrentiel de Birnbaum à trois paramètres (3-PLM). Le modèle 2-PLM a été conservé parce que l’ajustement n’a pas produit d’amélioration significative ; pour les items dichotomiques à CM, c’est le modèle 2-PLM qui a été utilisé. Les items polytomiques à RCS ont été calibrés à l’aide du modèle généralisé de crédit partiel (MGCP). Pour l’estimation de l’ensemble des trois paramètres des items, c’est la méthode *Maximum Marginal Likelihood Estimation – MMLE* (estimation de la probabilité marginale maximale) qui a été utilisée.

---

<sup>24</sup> Il vaut peut-être la peine de mentionner que le Système d’analyse statistique (SAS) a produit des résultats comparables.

<sup>25</sup> En 2010, la comparaison entre le niveau en lecture en 2007 et en 2010 s’est faite en utilisant comme valeurs de référence les paramètres des items de 2010 (voir CMEC, 2011). La décision d’utiliser 2010 comme année de référence au lieu de 2007 a été prise en raison du changement de population cible (les élèves de 8<sup>e</sup> année/2<sup>e</sup> 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.

### ***Évaluation du caractère convenable des modèles de la TRI***

L'évaluation 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  $\chi^2$  (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  $\chi^2$  normalisées qui deviennent élevées 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 configuration effective des données a été évaluée à partir des statistiques  $S-\chi^2$ . (IRTPRO ne produit pas et n'approuve pas la production de la courbe de réponse aux items empirique.) Les statistiques  $S-\chi^2$  se fondent sur la différence entre les fréquences observées et les fréquences attendues dans les catégories de réponses selon les scores additionnés.

### ***Fonctionnement différentiel des items***

Pour le fonctionnement différentiel des items (FDI), il a fallu évaluer dans quelle mesure certains des items du PPCE de 2013 présentaient des propriétés statistiques différentes (niveau de difficulté, etc.) selon le sexe et la langue. Cela s'est fait à l'aide de la méthode Mantel-Haenszel (M-H) et du test de Wald selon l'implémentation dans IRTPRO. Ce test se fait dans IRTPRO [*Traduction libre*] « avec des matrices exactes de variance-covariance des erreurs dans les paramètres des items calculées au moyen d'un algorithme espérance-maximisation complété » (voir la documentation technique du IRTPRO). Si certains des items ont bel et bien manifesté un fonctionnement différentiel, cela s'est fait de façon équilibrée entre les groupes comparés, comme l'a montré le chevauchement quasi parfait du fonctionnement différentiel du test.

### ***Liens et équivalences avec les évaluations précédentes pour les domaines secondaires***

Le travail d'établissement des liens et d'égalisation a donné une indication de l'évolution entre les évaluations précédentes et la dernière évaluation. Les domaines secondaires du PPCE de 2013 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 de 2013. Par conséquent, tous les items en lecture et en mathématiques étaient des items d'ancrage. La conception de l'évaluation correspondait à la méthode *Nonequivalent groups with Anchor Test* – NEAT (méthode des groupes non équivalents avec ancrage). Cependant, comme le changement de définition de la population cible a conduit à choisir l'année 2010 comme année de référence en lecture, les évaluations de 2013 en lecture et en mathématiques étaient toutes deux liées à l'évaluation de 2010. Ce lien s'est fait au moyen d'un calibrage simultané. Comme les paramètres de deux items

d'évaluation successifs étaient estimés simultanément avec des items communs (p. ex., les items en mathématiques de 2010 et de 2013), les paramètres des items communs avaient les mêmes estimations et se situaient sur la même échelle (de Ayala, 2009; Kim et Kolen, 2006). Cette approche avait l'avantage qu'elle faisait une utilisation maximale de toutes les données disponibles pour l'estimation des paramètres des items (Martin, Mullis, Foy, Brossman et Stanco, 2012).

Pour ce qui est de l'estimation des scores  $\theta$ , 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 simultané, 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 de 2013 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 simultané est fait, les étapes suivantes :

- estimer les distributions des niveaux de rendement pour l'évaluation actuelle en utilisant le paramètre du calibrage simultané ;
- déterminer la transformation linéaire qui fait le mieux concorder les distributions des niveaux de rendement de l'évaluation précédente (par exemple, du PPCE de 2010 en mathématiques) estimées dans le cadre du calibrage simultané aux mêmes distributions de l'évaluation obtenues lors de l'estimation initiale des paramètres des items (par exemple, les estimations utilisées en 2010) ;
- applique la transformation linéaire de l'étape II aux distributions des niveaux de rendement de l'évaluation actuelle (par exemple, l'évaluation du PPCE de 2013 en mathématiques).

La production des scores pour le niveau de rendement en lecture et en mathématiques pour le PPCE de 2013 (scores  $\theta$  et scores à l'échelle) a utilisé les paramètres des items estimés à ce stade.

### **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.

### **Production des scores pour le rendement et mise à l'échelle**

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

- Un score  $\theta$  a été produit pour représenter le niveau global de l'élève dans le domaine en question (sciences, lecture ou mathématiques). L'estimation du score  $\theta$  s'est faite à l'aide de la méthode *Expected A Posteriori* – EAP (attente a posteriori).

- La production des scores thêta en sciences s'est servie des paramètres des items du calibrage des items en sciences.
- Pour la lecture et les mathématiques, c'est le calibrage simultané établissant un lien entre les évaluations antérieures et le PPCE de 2013 qui a fourni les estimations des paramètres des items pour la production des scores thêta.
- Le score thêta a été rapporté à une échelle avec une moyenne nationale non pondérée de 500 et un écart-type de 100.
- Les scores à l'étape II ont été pondérés avec la pondération d'échantillonnage et la moyenne nationale été remise à 500 avec un écart-type de 100.

Les trois mêmes types de scores ont également été calculés pour tous les élèves dans chacun des quatre sous-domaines en sciences et dans les trois compétences ciblées.

### **Estimations des erreurs-types**

L'analyse des données du PPCE de 2010 s'est servie d'une approche *bootstrap* pour produire de façon empirique les estimations des erreurs-types pour les résultats nationaux et les moyennes par instance pour chacun des trois domaines évalués. Même si l'approche *bootstrap* est de plus en plus largement utilisée, en particulier dans les domaines de la recherche, elle peut être affectée par la production d'estimations qui ne sont pas cohérentes si la valeur de départ change à chaque exécution. Cela est dû au fait qu'il existe de nombreux échantillons aléatoires qui peuvent être tirés de l'échantillon initial. Du coup, si quelqu'un souhaite reproduire l'erreur-type pour les résultats du PPCE, il existe une certaine probabilité qu'il obtienne des résultats différents.

### **Présentation des résultats de l'évaluation du PPCE de 2013**

Des rapports sommaires sur les scores ont été élaborés au niveau national, au niveau des instances, selon la langue et selon le sexe pour chacun des trois domaines évalués. Les résultats ont été fournis sous forme de tableaux et de graphiques selon la présentation utilisée pour les rapports publics du PPCE de 2007 et de 2010.

## Chapitre 10 – Analyse des données du questionnaire

Comme lors des évaluations précédentes, le PPCE de 2013 a recueilli des informations contextuelles auprès des élèves, du personnel enseignant et des directions des écoles (à l'aide du questionnaire de l'école). L'analyse des données des questionnaires a compris les étapes suivantes :

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

Ces analyses statistiques ont été faites pour chacun des trois questionnaires et ont été présentées pour chaque langue.

### **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 la recodification 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.-à-d. 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**

Des tableaux de fréquence ont été produits pour chaque item :

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

### **Recodification des items**

Le PPCE de 2013 comprenait à la fois des réponses valables et des réponses non valables. La réponse à une question était valable si l'élève n'avait choisi qu'une option pour la réponse, mais elle était considérée comme non valable si l'élève avait choisi plus d'une réponse. La tâche

décrite ici a consisté à recodifier 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.

### ***Données manquantes***

Il y a eu trois types de données manquantes dans les questionnaires du PPCE de 2013 :

- réponses manquantes parce que l'élève n'avait pas eu assez de temps pour remplir le questionnaire (item non atteint)<sup>26</sup>;
- items omis, c'est-à-dire items sautés par l'élève volontairement ou non à des endroits quelconques dans le questionnaire;
- données manquantes parce que les questionnaires des enseignantes et enseignants ou des élèves remplis sur papier sont retournés après la capture des données s'est terminée.<sup>27</sup>

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

### **Analyse factorielle**

L'analyse factorielle a fait intervenir l'exécution d'une analyse factorielle exploratoire (AFE) du PPCE de 2013.

---

<sup>26</sup> Il n'y avait pas de limite de temps pour le personnel enseignant et pour les directions des écoles, de sorte que les résultats n'étaient pas censés avoir des données manquantes.

<sup>27</sup> Les données de 25 pour cent seulement des élèves et des enseignantes et enseignants du Nouveau-Brunswick sont capturées parce que les questionnaires remplis sur papier étaient retournés plusieurs mois après que la capture des données soit terminée. Seulement les données des questionnaires retournées en ligne par cette population sont incluses.

### **Analyse des items : analyse de la théorie classique des items**

Une analyse statistique classique a permis l'évaluation de certaines propriétés de mesure des items des questionnaires contextuels du PPCE de 2013. Elle a été effectuée pour chaque questionnaire et les résultats ont été présentés par instance et par langue. Elle s'est concentrée principalement sur les éléments suivants :

- moyenne et écart-type ;
- discrimination des items ;
- pourcentage de personnes interrogées ayant une réponse manquante pour l'item ;
- indice de cohérence interne pour chaque échelle dans chaque questionnaire.

### **Analyse de comparaison des groupes**

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

- comparer les moyennes de rendement ou certaines variables liées aux élèves (p. ex., l'attitude de l'élève vis-à-vis des sciences, la perception qu'il a de sa propre efficacité, son assurance) par rapport à une variable catégorique donnée (par exemple, le sexe, l'emplacement de l'école — en milieu rural ou urbain) tirée d'un des questionnaires (questionnaire de l'élève, questionnaire du personnel enseignant ou questionnaire de l'école) ;
- comparer le niveau de rendement par rapport à une variable catégorique donnée tirée d'un des questionnaires (questionnaire de l'élève, questionnaire du personnel enseignant ou questionnaire de l'école) ;
- comparer la moyenne de rendement par rapport à un ensemble de variables contextuelles catégoriques se rapportant à l'élève, au personnel enseignant ou à l'école ;
- comparer les moyennes de rendement (autres variables des élèves, comme l'attitude de l'élève vis-à-vis des sciences, la perception qu'il a de sa propre efficacité, son assurance) par rapport à un ensemble de variables catégoriques se rapportant au personnel enseignant ;
- comparer les moyennes de rendement (autres variables des élèves, comme l'attitude de l'élève vis-à-vis des sciences, la perception qu'il a de sa propre efficacité, son assurance) par rapport à un ensemble de variables catégoriques se rapportant à l'école.

### **Analyse corrélacionnelle**

L'analyse corrélacionnelle a fait intervenir les éléments suivants :

- calculer des coefficients de corrélation simple – aussi appelée corrélation bivariée ou corrélation d'ordre zéro – entre le niveau de rendement des élèves et une variable calculée (tirée d'un questionnaire) ou une variable contextuelle donnée sur une échelle. Comme les coefficients de corrélation étaient produits sous forme de matrices, ils comprenaient également la corrélation entre les variables calculées et entre les variables calculées et les autres variables contextuelles ;
- effectuer une analyse de régression linéaire multiple pour prédire le rendement en sciences à partir d'un ensemble de variables se rapportant aux élèves, notamment de variables catégoriques ;

- 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 sciences, ainsi que le rendement en lecture et en mathématiques, à partir d'un ensemble de variables se rapportant au personnel enseignant ou à l'école, notamment de variables catégoriques ;
- 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 sciences, ainsi que le rendement en lecture et en mathématiques, à partir d'un ensemble de variables se rapportant au personnel enseignant ou à l'école, notamment de variables catégoriques.

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

## Chapitre 11 – Base de données du PPCE

### Description des bases de données

Toutes les bases de données du PPCE de 2013 sont en français et en anglais et sont à la disposition des chercheuses et chercheurs. Le CMEC a plusieurs bases de données, dont l'une couvrant l'ensemble des élèves participants, l'une couvrant l'ensemble des écoles participantes et l'une couvrant l'ensemble des enseignantes et enseignants des élèves participants. Il y a également une base de données élèves/personnel enseignant/écoles qui contient tous les enregistrements des élèves fusionnés aux réponses aux questionnaires. Cette base de données permet d'établir des liens entre le rendement des élèves et les données contextuelles. Vous trouverez ci-dessous une description des bases de données SPSS et Excel<sup>28</sup>.

#### ***Base de données des élèves***

Cette base de données contient principalement les informations suivantes :

- renseignements généraux sur les élèves (numéros d'identification de l'élève, de l'école et de l'enseignante/enseignant ; code de participation de l'élève ; numéro du cahier d'évaluation ; province ou territoire et langue de l'élève) ;
- pondérations statistiques pour les élèves et les écoles ;
- réponses au questionnaire de l'élève ;
- scores factoriels associés aux élèves ;
- résultats du test de rendement (mathématiques, sciences et lecture).

#### ***Base de données du personnel enseignant***

Cette base de données principalement les informations suivantes :

- renseignements généraux sur le personnel enseignant (numéros d'identification de l'enseignante/enseignant et de l'école ; province ou territoire et langue pour chaque enseignante/enseignant) ;
- réponses au questionnaire du personnel enseignant ;
- scores factoriels associés au personnel enseignant.

Les chiffres définitifs permettant de déterminer le nombre d'enseignantes et enseignants ne sont pas disponibles, car l'échantillon d'enseignantes et enseignants était basé sur les échantillons des écoles et des élèves. Toutes les enseignantes et tous les enseignants ayant enseigné les sciences aux élèves effectuant le test du PPCE dans une école donnée ont été inclus dans l'échantillon ; il manque toutefois un grand nombre de numéros d'identification de l'enseignante/enseignant. Comme des classes intégrales ont été sélectionnées, une enseignante ou un enseignant seulement a fait l'objet d'un échantillonnage dans la plupart des écoles, avec deux ou plusieurs enseignantes ou enseignants dans quelques écoles.

---

<sup>28</sup> Des listes de codages qui identifient les variables et les valeurs dans le contenu des bases de données sont disponibles sur demande.

### ***Base de données des écoles***

Cette base de données contient principalement les informations suivantes :

- renseignements généraux sur les écoles (numéro d'identification de l'école, province ou territoire et langue de l'école) ;
- pondération statistique pour les écoles ;
- réponses des directions des écoles au questionnaire de l'école ;
- scores factoriels associés aux écoles.

### ***Base de données fusionnée — élèves/personnel enseignant/écoles***

Cette base de données contient 700 variables et 32 379 enregistrements. Elle contient toutes les informations des bases de données ci-dessus (base de données des élèves, base de données du personnel enseignant et base de données des écoles). Cette base de données permettra aux chercheuses et chercheurs d'établir des liens entre le rendement des élèves et les données contextuelles.

### **Consultation de la base de données pour les recherches**

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

Le CMEC s'engage à promouvoir la recherche ayant trait aux politiques 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/territorial. 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 instance donnée (p. ex., les écoles ou les commissions/conseils 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, 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 instances soient tenues 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 à la base de données :

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

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

Les demandes d'accès au matériel d'évaluation confidentiel (p. ex., les cahiers des épreuves) seront étudiées par le CMEC à la condition stricte que le contenu des cahiers 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 instances membres nommées ou identifiées dans le rapport, en vertu d'un engagement de confidentialité, un mois avant la publication ou la parution des conclusions, afin que les instances 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.

### **Coordonnées**

Pour en savoir plus sur les données du PPCE, veuillez communiquer avec [pcapinfo@cmecc.ca](mailto:pcapinfo@cmecc.ca).

## Bibliographie

- BAKER, F. B., et A. AL-KARNI. « A comparison of two procedures for computing IRT equating coefficients », *Journal of Educational Measurement*, n° 28, 1991, p. 147-162.
- BENJAMIN, R. « Reconstructing readability: Recent developments and recommendations in the analysis of text difficulty », *Educational Psychology Review*, vol. 24, n° 1, 2012, p. 63-88.
- BRIGGS, D. « An introduction to Multidimensional IRT », communication présentée à UC Berkeley, avril 2008. Sur Internet :  
[http://www.powershow.com/view/3c4039-MmRjY/An\\_Introduction\\_to\\_Multidimensional\\_IRT\\_Derek\\_Briggs\\_April\\_powerpoint\\_ppt\\_presentation](http://www.powershow.com/view/3c4039-MmRjY/An_Introduction_to_Multidimensional_IRT_Derek_Briggs_April_powerpoint_ppt_presentation)
- CAI, L. « Metropolis-Hastings Robbins-Monro algorithm for confirmatory item factor analysis », *Journal of Educational and Behavioral Statistics*, vol. 35, n° 3, 2010, p. 307-335.
- CHEN, W.-H., et D. THISSEN. « Local dependence indexes for item pairs using item response theory », *Journal of Educational and Behavioral Statistics*, vol. 22, n° 3, 1997, p. 265-289.
- CIZEK, G. J., et M. B. BUNCH. *Standard setting: A guide to establishing and evaluating performance standards on tests*, Thousand Oaks, CA, SAGE Publications Ltd., 2007.
- CONSEIL DES MINISTRES DE L'ÉDUCATION (CANADA) [CMEC]. *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) [CMEC]. *Étude documentaire des programmes d'études des sciences et des cadres d'évaluation*, Toronto, CMEC, 2005 [inédit].
- CONSEIL DES MINISTRES DE L'ÉDUCATION (CANADA) [CMEC]. *PPCE de 2010 – Programme pancanadien d'évaluation*, Toronto, CMEC, 2011.
- DE AYALA, R. J. *The theory and practice of item response theory*, New York, Guilford Press, 2009.
- DE AYALA, R. J., B. S. PLAKE et J. C. IMPARA. « The impact of omitted response on the accuracy of ability estimation in item response theory », *Journal of Educational Measurement*, vol. 38, n° 3, 2001, p. 213-234.
- DE CHAMPLAIN, A. F., et M. E. GESSAROLI. « Assessing the dimensionality of item response matrices with small sample size and short test lengths », *Applied Measurement in Education*, n° 11, 1998, p. 231-253.
- FENSHAM, P., et W. HARLEN. « School science and public understanding of science », *International Journal of Science Education*, vol. 21, n° 7, 1999, p. 755-763.
- FINCH, H. « Estimation of item response theory parameters in the presence of missing data », *Journal of Educational Measurement*, vol. 45, n° 3, p. 225-245.

- FINCH, H., et B. HABING. « Comparison of NOHARM and DETECT in item cluster recovery: Counting dimension and allocating items », *Journal of Educational Measurement*, n° 42, p. 149-169.
- FRASER, C., et R. P. McDONALD. *NOHARM: A window program for fitting both unidimensional and multidimensional normal ogive models of latent trait theory* [logiciel], Welland (Ontario), Niagara College, 2003. Sur Internet : <http://noharm.software.informer.com>
- FOY, P., B. BROSSMAN et J. GALIA. « Scaling the TIMSS and PIRLS 2011 achievement data », dans M. O. Martin et I. V. S. Mullis (éd.), *Methods and procedures in TIMSS and PIRLS 2011*, Chestnut Hill, MA, TIMSS & PIRLS International Study Center, Boston College, 2012. Sur Internet : [http://timssandpirls.bc.edu/methods/pdf/TP11\\_Scaling\\_Achievement.pdf](http://timssandpirls.bc.edu/methods/pdf/TP11_Scaling_Achievement.pdf)
- HIDI, S., et D. BERNDORFF. « Situational interest and learning », dans L. Hoffmann, A. Krapp, K.A. Renniger et J. Baumert (éd.), *Interest and Learning*, Kiel (Allemagne), Institut des sciences de l'éducation de l'Université de Kiel, 1998.
- HOY, A. W. « Changes in teacher efficacy during the early years of teaching », communication présentée à l'assemblée générale annuelle de l'*American Educational Research Association* (association américaine de la recherche sur l'éducation), Nouvelle-Orléans, avril 2000.
- JOHNSON, M. S. « Marginal maximum likelihood estimation of item response models », *Journal of Statistical Software*, vol. 20, n° 10, 2007, p. 1-19.
- KANDEL, L., et A. MOLES. « Application de l'indice de Flesch à la langue française », *Cahiers d'études de radio-télévision*, n° 19, 1958, p. 253-274.
- KIM, S., et M. J. KOLEN. « Robustness to format effects of IRT linking methods for mixed-format tests », *Applied Psychological Measurement*, vol. 19, n° 4, 2006, p. 357-381.
- KLARE, G. R. « The formative years », dans B. L. Zakaluk et S. J. Samuels (éd.), *Readability, its past, present and future*, Newark, Delaware, International Reading Association, 1988, p. 14-34.
- KNOL, W. R., et M. P. F. BERGER. « Empirical comparison between factor analysis and multidimensional item response models », *Multivariate Behavioral Research*, n° 26, 1991, p. 457-477.
- LAVEAULT, D., et J. GRÉGOIRE. *Introduction aux théories des tests en psychologie et en éducation*, 2<sup>e</sup> éd., Bruxelles, De Boeck, 2002.
- LORD, F. M. « Estimation of latent ability and item parameters when there are omitted responses », *Psychometrika*, n° 39, 1974, p. 247-264.
- LORD, F. M. « Maximum likelihood estimation of item parameters when some responses are omitted », *Psychometrika*, n° 48, 1983, p. 477-482.
- LUDLOW, L. H., et M. O'LEARY. « Scoring omitted and not-reached items: Practical data analysis implications », *Educational and Psychological Measurement*, n° 59, 1999, p. 615-630.

- McDONALD, R. P. *Nonlinear factor analysis*, coll. « Psychometric Monographs », n° 15, Richmond, VA, Psychometric Corporation, 1967. Sur Internet :  
<https://www.psychometricsociety.org/sites/default/files/pdf/MN15.pdf>
- MARTIN, M. O., I. V. S. MULLIS, P. FOY, B. BROSSMAN et G. M. STANCO. « Estimating linking error in PIRLS », *IERI Monograph Series: Issues and Methodologies in Large-Scale Assessments*, n° 5, 2012, p. 35-47. Sur Internet :  
[http://www.ierinstitute.org/fileadmin/Documents/IERI\\_Monograph/IERI\\_Monograph\\_Volume\\_05\\_Chapter\\_2.pdf](http://www.ierinstitute.org/fileadmin/Documents/IERI_Monograph/IERI_Monograph_Volume_05_Chapter_2.pdf)
- MURAKI, E., et G. ENGELHARD. « Examining differential item functioning with BIMAIN », communication présentée à l'assemblée générale annuelle de l'American Educational Research Association, San Francisco, avril 1989.
- NATIONAL COUNCIL OF TEACHERS OF MATHEMATICS. *Principles and Standards for School Mathematics*, Reston, VA, NCTM, 2000.
- ORGANISATION DE COOPÉRATION ET DE DÉVELOPPEMENT ÉCONOMIQUES (OCDE). *PISA 2009 Technical Report*, Paris, PISA, Éditions OCDE, 2012. Sur Internet :  
<http://dx.doi.org/10.1787/9789264167872-en>
- ORGANISATION DE COOPÉRATION ET DE DÉVELOPPEMENT ÉCONOMIQUES (OCDE). *PISA 2006 – Les compétences en sciences, un atout pour réussir*, Paris, PISA, Éditions OCDE, 2006.
- OSBORNE, J., S. SIMON et S. COLLINS. « Attitudes towards science: A review of the literature and its implications », *International Journal of Science Education*, vol. 25, n° 9, 2003, p. 1049-1079.
- RIGGS, I., et L. ENOCHS. « Towards the development of an elementary teacher's science teaching efficacy belief instrument », *Science Education*, n° 74, 1990, p. 625-637.
- WANG, M. C., G. D. HAERTEL et H. J. WALBERG. « Synthesis of research: What Helps Students Learn? », *Educational Leadership*, décembre 1993/janvier 1994, p. 74-79.
- WANG, M. C., G. D. HAERTEL et H. J. WALBERG. « Toward a knowledge base for school learning », *Review of Educational Research*, vol. 63, n° 3, 1993, p. 249-294.
- WANG, M. C., G. D. HAERTEL et H. J. WALBERG. « What influences learning? A content analysis of review literature », *Journal of Educational Research*, vol. 84, n° 1, 1990, p. 30-43.
- YEN, W. M. « Scaling performance assessments: Strategies for managing local item dependence », *Journal of Educational Measurement*, n° 30, 1993 p. 187-213.
- ZHANG, B., et C. M. WALTER. « Impact of missing data on person-model fit and person trait estimation », *Applied Psychological Measurement*, vol. 32, n° 6, 2008, p. 466-479.