

PCAP 2010 - RAPPORT TECHNIQUE
PARTIE I
Chapitre 3

MISE À L'ESSAI DES ITEMS

Table des matières

1. Cahiers d'évaluation.....	2
2. Séance de correction des items	2
3. Capture de données	3
4. Analyse des données	3
5. Séance de révision des items	4
6. Révision du cadre d'évaluation.....	5

MISE À L'ESSAI DES ITEMS

Les items présentés aux élèves dans les évaluations du rendement 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'enquête réelle afin que seuls les meilleurs soient retenus pour évaluer le rendement des élèves de 8^e année.

1. Cahiers d'évaluation

La mise à l'essai des items du PPCE 2010 a eu lieu au printemps 2009 pour les mathématiques et les sciences. Approximativement 2000 élèves francophones et anglophones d'environ 100 écoles de toutes les instances ont répondu aux items en mathématiques et en sciences. Le nombre d'items en mathématiques surpassait celui des items en sciences puisque les mathématiques étaient le principal domaine de l'évaluation de 2010. Les élèves avaient 60 minutes pour compléter l'évaluation et 30 minutes pour répondre aux questions du Questionnaire de l'élève. La mise à l'essai a permis de tester de nouveaux items et de vérifier s'ils fonctionnaient bien pour les élèves de 8^e année.

En tout, quatre cahiers d'évaluation (en anglais et en français) ont été présentés aux élèves, et ils comprenaient tous des items en mathématiques et en sciences. Dans chaque cahier, deux blocs, l'un au début et l'autre à la fin du cahier, étaient réservés pour les mathématiques alors que l'autre bloc, au milieu du cahier, était réservé pour les sciences. Le dernier bloc du cahier (réservé aux mathématiques) se trouvait au début du cahier suivant. L'intention était de connaître la manière dont les élèves répondaient aux items en comparant les réponses des élèves ayant répondu au début du cahier à celles des élèves ayant répondu à la fin du cahier. Deux des quatre cahiers comprenaient chacun huit scénarios (ou sections), et les deux autres cahiers en comprenaient chacun sept. Chaque cahier d'évaluation incluait à la fois des items à réponse choisie et des items à réponse construite. Deux cahiers d'évaluation comprenaient 38 items chacun, alors que les deux autres comprenaient 39 items chacun. Pour les sciences, l'évaluation comprenait aussi des items d'attitude pour mesurer les attitudes des élèves envers l'environnement. Ces items n'ont toutefois pas été pris en compte dans la note globale. Les élèves francophones et anglophones ont répondu aux mêmes items dans l'évaluation de la mise à l'essai. Il n'y a pas eu de mise à l'essai pour la lecture étant donné qu'il s'agissait du domaine principal en 2007, et que seulement certains items de l'évaluation principale de cette même année seraient réutilisés pour l'évaluation de 2010. Il y avait assez de données statistiques sur les items en lecture étant donné qu'ils avaient été présentés à environ 30 000 élèves en anglais et en français en 2007.

2. Séance de correction des items

La séance de correction des items en mathématiques et en sciences a eu lieu à Gatineau, du 6 au 17 juillet 2009. La correction des mathématiques devait être faite en 10 jours alors que

la correction des sciences devait se faire au cours des trois derniers jours. Approximativement 2000 cahiers ont été corrigés, dont environ 1000 cahiers en anglais et 1000 cahiers en français. La séance comportait deux chefs de tables en mathématiques, dont l'un pour la table anglophone et l'autre pour la table francophone, ainsi que 14 correctrices et correcteurs, dont la moitié se trouvait à la table francophone et l'autre moitié à la table anglophone. Pour les sciences, il y avait également deux chefs de tables, dont l'un pour la table anglophone et l'autre pour la table francophone, ainsi que 10 correctrices et correcteurs, dont la moitié corrigeait les items anglophones et l'autre moitié les items francophones.

Le processus de correction employait 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.

3. Capture de données

Les élèves ont écrit leurs réponses directement dans les cahiers d'évaluation du PPCE en encerclant la réponse choisie (pour les items à réponse choisie) ou en écrivant quelques lignes (pour les items à réponse construite).

Lors de la séance de correction pour la mise à l'essai, les correctrices et correcteurs avaient reçu chacun une liasse de 10 cahiers d'évaluation qui provenaient de différentes instances. Ils ont corrigé seulement les items à réponse construite en remplissant le cercle d'une feuille de code qui correspondait le mieux à la réponse de l'élève. Une fois la séance de correction terminée, tous les cahiers d'évaluation ont été envoyés à une entreprise 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, de même que 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.

4. Analyse des données

Le CMEC a fait appel à un groupe de spécialistes en statistiques en éducation pour effectuer l'analyse des données de la mise à l'essai pour les items en mathématiques et en sciences. Ce groupe de spécialistes était chargé de préparer les bases de données, les fichiers et les règles pour la capture de données ainsi que de valider les données pour les deux matières évaluées.

Une seule base de données pour la mise à l'essai a été créée, incluant à la fois les items en mathématiques et en sciences. La base de données incluait également des données sur le questionnaire de même que, pour chaque élève, une note globale, une note pour chaque scénario et une note pour chaque sous-domaine.

Le CMEC et le groupe de spécialistes ont aussi établi des règles pour la capture de données, c'est-à-dire pour la pondération des items. Les items en mathématiques et en sciences

incluaient à la fois des items à réponse choisie et des items à réponse construite, ces derniers offrant la possibilité d'octroi de crédit partiel. Il a donc été convenu d'attribuer un point aux items à réponse choisie par réponse exacte, et d'attribuer des valeurs selon une échelle de 0 à 3 pour les réponses à crédit partiel, la plus haute valeur possible étant 3.

L'analyse des données s'est effectuée selon la théorie classique, et les données obtenues ont été utilisées par les groupes de travail pour sélectionner les meilleurs items pour l'évaluation principale et pour rejeter les items aberrants, c'est-à-dire ceux qui ne se comportaient pas comme les autres items du test. Pour analyser les items, des indices statistiques ont été fournis aux groupes de travail, 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 sur le 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 différencie les élèves les plus forts des élèves les plus faibles. Il a aussi été nécessaire d'estimer la fidélité interne du test à l'aide du coefficient alpha de Cronbach.

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 qui ont sélectionné chaque réponse pour les items à réponse choisie ou qui ont obtenu chaque code pour les items à réponse construite. Des analyses utilisant le pourcentage des données manquantes ont également été réalisées.

5. Séance de révision des items

Après la mise à l'essai des items, les groupes de travail en mathématiques et en sciences se sont à nouveau réunis pour réviser les scénarios et les items et sélectionner ceux qui feraient partie de l'évaluation principale. Tous les cahiers d'évaluation leur avaient été fournis, de même que les résultats et les statistiques pour chaque item, afin qu'ils puissent vérifier la qualité et le degré de difficulté des items. 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 la qualité.

Le groupe de travail pour chaque matière 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.

Il est important de noter que les spécialistes en lecture se sont aussi réunis et ont sélectionné les items en se basant sur les données statistiques recueillies à partir de l'évaluation principale de 2007.

Le groupe de spécialistes a également révisé 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. Les spécialistes ont sélectionné les items qui, selon eux, pouvaient fournir le plus de données pertinentes lors de l'évaluation principale, comme l'établissement de liens entre les données contextuelles et le rendement des élèves.

6. Révision du cadre d'évaluation

La mise à l'essai des items a fourni de l'information sur les items qui a contribué à sélectionner les meilleurs items pour l'évaluation principale. Les spécialistes de contenu ont par la suite révisé le cadre d'évaluation afin d'y apporter les changements requis. Très peu de changements ont été apportés au cadre d'évaluation du PPCE 2010.