

Résumé du Rapport technique de l'OQRE

pour les tests de 2007-2008

*Tests en lecture, écriture et mathématiques,
cycle primaire (de la 1^{re} à la 3^e année) et cycle moyen (de la 4^e à la 6^e année),
Test de mathématiques, 9^e année, et
Test provincial de compétences linguistiques*



Introduction

En vue de remplir son mandat, l'OQRE effectue quatre tests annuels à l'échelle de la province auprès des élèves de langue française et de langue anglaise.

Les évaluations à grande échelle sont conçues de manière à mesurer le rendement des élèves à un certain moment, dans des conditions standardisées.

Les devis des tests de l'OQRE servent à élaborer les items à choix multiple et à réponse construite pour chaque test, afin que le test présente les mêmes caractéristiques chaque année.

Les tests à grande échelle de l'OQRE évaluent dans quelle mesure les élèves satisfont aux attentes et contenus d'apprentissage sélectionnés qui sont décrits dans *Le curriculum de l'Ontario*. Les tests contiennent des items à réponse construite qui demandent une réponse écrite ainsi que des items à choix multiple. Nous élaborons un document-cadre pour chaque test afin de décrire de manière détaillée de quelle façon le test sera élaboré.

Après leur administration, les tests produisent des résultats au niveau des élèves, des écoles, des conseils scolaires et de la province. Les résultats des tests procurent des renseignements utiles qui appuient la planification de l'amélioration par les écoles, les conseils scolaires et le ministère de l'Éducation de l'Ontario.

Ce résumé du rapport technique a un double objectif : décrire la qualité technique des tests de l'OQRE de l'année 2008 et préciser certaines des procédures et l'expertise professionnelle qui ont servi à assurer l'exactitude, la validité, la fidélité et l'intégrité psychométrique des tests suivants, administrés durant l'année scolaire 2007-2008 :

- Tests en lecture, écriture et mathématiques, cycle primaire (de la 1^{re} à la 3^e année) et cycle moyen (de la 4^e à la 6^e année);
- Test de mathématiques, 9^e année;
- Test provincial de compétences linguistiques (TPCL).

Méthodes utilisées pour produire les tests de l'OQRE

Élaboration des items

Il est essentiel de veiller à ce qu'on tienne compte des mêmes attentes et contenus d'apprentissage du curriculum de l'Ontario et que le degré de difficulté des tests soit similaire chaque année. À cette fin, tous les nouveaux items sont fondés sur une version détaillée des devis. L'OQRE recrute et forme des éducatrices et éducateurs spécialisés dans les domaines de la littérature (lecture et écriture en français et en anglais) et des mathématiques, venant de tous les coins de la province, afin d'appliquer les devis à leur travail dans le cadre des comités d'élaboration des items.

Chaque année, de nouveaux items sont élaborés et mis à l'essai pour devenir des items de la composante test dans le test de l'année suivante.

Nous mettons un point d'honneur à élaborer des items qui évaluent les attentes et contenus d'apprentissage du curriculum et qui soient appropriés, équitables et accessibles au plus grand nombre d'élèves de l'Ontario.

Les items mis à l'essai sont intégrés aux tests provinciaux de chaque année.

Avant de pouvoir être inclus dans un test, un item doit respecter des critères précis, être mis à l'essai, être examiné par des comités de contenu et de sensibilité ainsi qu'être révisé et édité. En ce qui concerne les items à choix multiple, les équipes examinent la clarté et le degré de complétude de l'amorce, l'exactitude de la réponse saisie et la plausibilité des trois leurres. En ce qui concerne les items de lecture et de mathématiques à réponse construite ainsi que les tâches d'écriture, les équipes étudient la correspondance des items avec leurs grilles de notation pour déterminer si les items vont obtenir la gamme de réponses attendues.

Le comité d'évaluation du contenu examine les items, pour s'assurer que ceux-ci sont appropriés à l'âge et à l'année d'études des élèves, les attentes et contenus d'apprentissage du curriculum qui sont évalués ainsi que le but du test. Le comité de sensibilité s'intéresse aux questions d'équité et d'accessibilité pour le plus grand nombre d'élèves en Ontario en examinant les questions d'équité en éducation (p. ex., questions relatives au sexe, au multiculturalisme, aux élèves inscrits à un programme d'ALF/de PDF et aux élèves ayant des besoins particuliers), afin de s'assurer qu'aucun groupe particulier d'élèves ne soit avantagé ou désavantagé par un des items figurant dans un test.

Mise à l'essai

Étant donné que les items de la composante mise à l'essai sont élaborés de la même manière que les items de la composante test, les élèves ne savent pas à quel type d'item ils répondent. Cette technique permet de circonvenir les élèves qui ont tendance à manquer de motivation lorsqu'ils répondent à de nouveaux items mis à l'essai en dehors du test provincial. Les résultats obtenus pour les items mis à l'essai ne sont pas utilisés pour établir les résultats des élèves, des écoles, des conseils scolaires ou de la province.

La mise à l'essai des éléments du test permet d'assurer que les items sélectionnés pour de futurs tests provinciaux sont satisfaisants et équitables sur le plan psychométrique pour tous les élèves. La mise à l'essai permet aussi de fournir des données qui sont utilisées pour mettre en équivalence le test de chaque année avec celui de l'année précédente de façon à pouvoir effectuer

une comparaison valide des résultats. La qualité de la comparaison des données d'une année à l'autre dépend de cette mise en équivalence.

Élaboration des tests

Il faut suivre un certain nombre de directives lors de l'élaboration des tests.

L'OQRE tient compte des directives ci-dessous pour sélectionner les items qui figureront dans ses tests provinciaux :

- Les items doivent être équitables et le degré de difficulté du test doit être comparable à celui des tests précédents.
- Les items doivent être conformes au devis et être équilibrés en ce qui concerne le contenu de la matière, la représentation des filles et des garçons ainsi que les aspects de la démographie provinciale (urbain/rural, nord/sud).
- Étant donné que les items se rapportent à des groupes d'attentes et contenus d'apprentissage du curriculum et que les attentes et contenus d'apprentissage d'un groupe ne seront pas tous évalués dans un test, il est important que l'ensemble des attentes et contenus d'apprentissage mesurables d'un groupe soient inclus dans au moins un test au cours d'un certain nombre d'années.

Questionnaires

Les informations contextuelles servent à analyser les données par groupe et à mettre l'accent sur les recommandations portant sur l'amélioration de l'apprentissage des élèves.

L'OQRE élabore des questionnaires à l'intention des élèves, du personnel enseignant et des directions d'école pour recueillir des informations contextuelles ainsi que des données sur les pratiques d'enseignement et les attitudes liées à la littératie et aux mathématiques. Les questions portent sur un certain nombre de facteurs liés au rendement des élèves.

Les guides d'administration décrivent les procédures garantissant que les tests sont administrés de façon uniforme et équitable à tous les élèves.

Administration des tests

Pour chaque test, il existe un guide d'administration qui décrit les rôles et responsabilités des enseignantes et enseignants et qui donne des directives sur la façon de préparer les élèves au test, sur ce qu'il faut leur dire ainsi que sur leur participation (c.-à-d. adaptations pour les élèves ayant des besoins particuliers, dispositions particulières pour les élèves inscrits à un programme d'ALF/de PDF, exemptions et reports).

Une autre façon pour l'OQRE d'assurer l'intégrité des procédures d'évaluation est d'affecter des observatrices et observateurs dans des écoles choisies de façon aléatoire durant les tests. En outre, l'organisme fait une analyse statistique des fichiers de données des réponses d'élèves pour chercher des types de réponses inhabituels. S'il trouve des types de réponses inhabituels, il communique avec l'école pour obtenir des renseignements additionnels.

Notation

Des procédures de notation rigoureuses assurent la fidélité des résultats des tests.

Les items à choix multiple sont notés de façon électronique. Les réponses rédigées aux items à réponse construite en lecture et en mathématiques et aux tâches d'écriture sont notées soigneusement et systématiquement par des personnes possédant les compétences requises, lesquelles sont principalement des éducatrices et éducateurs de l'Ontario. Une grille de notation générique pour chaque type d'item à réponse construite décrit le travail correspondant aux différents niveaux de rendement et maintient une uniformité entre les items et d'une année à l'autre. Une grille de notation spécifique aux items pour chaque item à réponse construite, pour chaque matière, est fondée sur les grilles de notation génériques. Les copies types illustrent les descripteurs pour les points bruts de la grille de notation.

Sélection des copies types

L'objectif de la sélection des copies types, qui est la première étape du processus de notation, est de définir et d'illustrer les divers niveaux de rendement pour chaque code des grilles de notation. Cette tâche est accomplie par un comité d'éducatrices et d'éducateurs ayant une expertise en notation et dans le système éducatif de l'Ontario. Les membres du comité s'entendent sur les réponses d'élèves qui représentent chaque point brut d'une grille de notation. Cet encodage des réponses d'élèves sert à sélectionner les copies types pour la grille de notation et à former les correctrices et correcteurs qui noteront les réponses des élèves.

Formation

Une formation poussée et des procédures de contrôle produisent des résultats présentant le plus haut niveau de validité.

Chaque correctrice ou correcteur suit une formation poussée afin de bien comprendre en quoi consiste le matériel de notation, afin que chaque chef de notation, son adjointe ou adjoint et chaque correctrice ou correcteur interprète et utilise le matériel de notation de la même façon.

Voici d'autres garanties et procédures de contrôle de la qualité :

- tests de qualification pour les correctrices et correcteurs, les chefs de notation et leurs adjointes et adjoints;
- supervision des correctrices et correcteurs par les chefs de notation et leurs adjointes et adjoints;
- examen quotidien et activités de calibrage;
- double notation anonyme de toutes les réponses des élèves qui ont fait le TPCL;
- procédures pour signaler des indications d'élèves à risque, un contenu inapproprié et l'intervention de l'enseignante ou de l'enseignant;
- contrôle quotidien en se référant à des copies de validation pour éviter toute dérive dans la notation;
- surveillance du taux d'exactitude des correctrices et correcteurs;
- normes de productivité;
- vérification des incohérences dans la notation;
- normes de rendement pour les chefs de notation et leurs adjointes et adjoints.

Toutes ces mesures visent à assurer que toute réponse d'élève soit notée de la même manière, indépendamment de la personne qui effectue la notation.

Mesure de la fidélité et de la validité de la notation

Les mesures de fidélité interjuges et de validité de notation servent à mesurer l'uniformité du rendement des correctrices et correcteurs. Nous comparons les résultats attribués à une réponse d'élève donnée pour voir s'ils sont exactement les mêmes, séparés d'un point brut (adjacents) ou de plus d'un point brut (non adjacents). Au moins deux correctrices et correcteurs notent chaque réponse d'élève pour le TPCL. Si les deux résultats ne sont pas identiques ou adjacents, une correctrice ou un correcteur spécialiste va trancher en ce qui a trait au résultat.

Fidélité interjuges

La fidélité interjuges est déterminée en comparant les résultats attribués par des pairs de correctrices et correcteurs à des réponses d'élèves choisies de façon aléatoire.

Pour la fidélité et la validité, l'OQRE a établi les normes industrielles ci-dessous :

- **concordance exacte de 60 % pour les grilles de notation à six points;**
- **concordance exacte de 65 % pour les grilles de notation à cinq points;**
- **concordance exacte de 70 % pour les grilles de notation à quatre points;**
- **concordance exacte de 75 % pour les grilles de notation à trois points;**

Pour les tests des cycles primaire et moyen, l'objectif de l'OQRE en matière de concordance exacte a été atteint pour 90 % des items de lecture en français et 50 % des items de lecture en anglais, pour 100 % des grilles de notation pour l'écriture en français et 42 % des grilles de notation pour l'écriture en anglais ainsi que pour 100 % des items de mathématiques en français et en anglais. L'objectif de concordance exacte et adjacente de l'OQRE a été atteint pour tous les items de lecture et toutes les tâches d'écriture en français et en anglais ainsi que pour tous les items de mathématiques en anglais et pour tous – à l'exception d'un – les items de mathématiques en français.

Pour le Test de mathématiques, 9^e année, les pourcentages de concordance exacte et de concordance exacte et adjacente ont dépassé les objectifs de l'OQRE pour tous les items dans les deux langues officielles.

Les pourcentages de concordance exacte n'ont atteint l'objectif de l'OQRE pour aucun des items de lecture du TPCL. Les pourcentages de concordance exacte et adjacente ont atteint l'objectif pour tous les items de lecture du TPCL. Pour les tâches d'écriture du TPCL, les pourcentages de concordance exacte ont atteint l'objectif uniquement pour les tâches d'écriture courte en anglais. Les pourcentages de concordance exacte et adjacente ont

- concordance exacte et adjacente de 95 % pour toutes les grilles.

La validité de la notation est établie en comparant, pour des réponses d'élèves choisies, les résultats attribués par les correctrices et correcteurs et ceux attribués par un panel d'experts.

atteint l'objectif pour 67 % des grilles de notation pour l'écriture en français et en anglais.

Validité de la notation

Pour les tests des cycles primaire et moyen, l'objectif de l'OQRE en matière de concordance exacte a été atteint pour 40 % des items de lecture en français et 15 % des items de lecture en anglais, pour 33 % des grilles de notation pour l'écriture en français et 8 % des grilles de notation pour l'écriture en anglais ainsi que pour 75 % des items de mathématiques en français et 94 % des items de mathématiques en anglais. L'objectif de l'OQRE pour une concordance exacte et adjacente a été atteint pour 70 % des items de lecture en français et tous les items de lecture en anglais, pour toutes les grilles de notation pour l'écriture ainsi que pour tous les items de mathématiques.

Pour le Test de mathématiques, 9^e année, les pourcentages de concordance exacte ont atteint l'objectif de l'OQRE pour tous les items en français et pour tous – à l'exception d'un – les items en anglais. Les pourcentages de concordance exacte et adjacente ont atteint l'objectif pour un item en français et pour tous – à l'exception d'un – les items en anglais.

Pour le TPCL, les pourcentages de concordance exacte n'ont atteint l'objectif de l'OQRE pour aucun des items de lecture en français et un seul item de lecture en anglais. Les pourcentages de concordance exacte ont atteint l'objectif pour la grille de notation des conventions linguistiques pour une tâche d'écriture longue en français et pour les tâches d'écriture courte en anglais. Les pourcentages de concordance exacte et adjacente ont atteint l'objectif pour tous les items de lecture en français et pour tous – à l'exception d'un – les items en anglais. Les pourcentages de concordance exacte et adjacente ont atteint l'objectif pour toutes les grilles de notation pour l'écriture à l'exception des deux grilles de conventions linguistiques en français et d'une grille de contenu en anglais, pour la tâche d'écriture longue.

Lorsque les correctrices et correcteurs n'étaient pas d'accord avec les résultats attribués par le panel, ils avaient la même tendance à attribuer des résultats plus élevés et plus bas que ceux du panel pour la majorité des composantes du test. La plus grande différence concernait la composante écriture du test du cycle moyen en français, pour laquelle 31 % des résultats attribués étaient adjacents quoique plus élevés que ceux attribués par le panel d'experts et 7 % des résultats étaient adjacents, mais plus bas que ceux attribués par le panel d'experts. Le pourcentage de résultats adjacents et bas était un peu plus haut que le pourcentage de résultats adjacents et élevés pour les composantes suivantes : lecture du test du cycle primaire et mathématiques du test du cycle moyen en français; écriture et mathématiques du test du cycle primaire en anglais. Le pourcentage de résultats adjacents et élevés était un peu plus haut que le pourcentage de résultats adjacents et bas pour l'écriture et les mathématiques du test du cycle primaire en français.

Notation de la mise à l'essai

La notation d'items à réponse construite de la composante mise à l'essai se fait après la notation des items de la composante test. Les correctrices et correcteurs les plus fiables et les plus productifs en notation de la composante test effectuent la notation des items de la composante mise à l'essai afin d'assurer l'uniformité d'une année à l'autre et de réduire le temps requis pour former les correctrices et correcteurs. Le personnel de notation de la mise à l'essai est formé à l'aide d'items mis à l'essai et de tâches, de grilles de notation spécifiques aux items et d'exigences en matière de notation, afin de produire des données spécifiques aux items et aux tâches qui sont valides et fidèles en vue de la conception des tests provinciaux et de la mise en équivalence.

Mise en équivalence et analyse des résultats des élèves

La mise en équivalence est un processus qui place les résultats des élèves de plusieurs années consécutives sur une même échelle afin que les résultats de rendement des élèves puissent être comparés de façon valide au fil des ans.

La théorie de la réponse à l'item (TRI) définit la relation qui existe entre la capacité de l'élève, relativement à un construit déterminé, et la probabilité que l'élève réponde correctement à un item qui évalue ce construit.

Différents modèles de la TRI sont utilisés pour les tests des cycles primaire et moyen, le test de 9^e année et le TPCL.

L'OQRE utilise divers modèles, conceptions et analyses statistiques, afin d'assurer la comparabilité des résultats des tests d'une administration à l'autre. Le processus de mise en équivalence met les résultats d'élèves de deux années adjacentes sur une même échelle. Cela permet de contrôler les petites différences dans le degré de difficulté des tests d'une année à l'autre, de faire des comparaisons valables d'une année à l'autre et de veiller à ce que les élèves d'une année donnée ne soient pas avantagés au détriment des élèves d'une autre année.

Les seuils de rendement sont ensuite déterminés pour indiquer les niveaux de rendement utilisés dans la communication des résultats de l'OQRE (quatre niveaux de rendement pour le cycle primaire, le cycle moyen et la 9^e année et réussite ou non réussite pour le TPCL). Les élèves qui reçoivent un résultat donné dans l'année en cours auront démontré le même niveau d'habileté et de connaissances que les élèves des années antérieures auxquels on a attribué ce même niveau.

L'OQRE utilise la TRI pour le calibrage des items, la notation et la mise en équivalence. Les modèles généraux de la TRI utilisés par l'OQRE sont un modèle logistique modifié à deux paramètres pour les tests des cycles primaire et moyen et de 9^e année et un modèle de Rasch modifié pour le TPCL. Les deux modèles sont modifiés en ajoutant un paramètre de pseudo-chance fixé à 0,20 pour les items à choix multiple. L'OQRE se sert du modèle à crédit partiel généralisé pour les items à réponse construite de tous les tests.

Le plan du groupe non équivalent d'items communs à paramètre fixe-avant a été utilisé pour mettre en équivalence les tests de l'OQRE administrés en 2007-2008 et en 2006-2007. Les tests provinciaux de l'année scolaire 2007-2008 ont été créés à partir d'items mis à l'essai en 2006-2007 et ces items ont servi à établir un lien entre les tests. Les tests provinciaux de 2007-2008 ont été calibrés et les paramètres des items ont été fixés sur les items qui avaient été mis à l'essai dans les tests de 2006-2007. Les tests provinciaux de 2006-2007 ont

ensuite été calibrés avec les items mis à l'essai afin de mettre les tests de 2006-2007 et de 2007-2008 sur une même échelle. Les paramètres des items de ce calibrage ont ensuite été utilisés pour noter de nouveau les réponses d'élèves de l'année scolaire 2006-2007. Enfin, les valeurs θ correspondant au pourcentage d'élèves à chaque niveau de rendement en 2006-2007 ont été déterminées, puis appliquées aux distributions de l'année scolaire 2007-2008 pour déterminer le pourcentage d'élèves à chaque niveau de rendement en 2007-2008.

Les procédures de mise en équivalence et la détermination des seuils de rendement ont été réalisées séparément pour les versions française et anglaise des tests.

Toutes les analyses ont été faites à deux reprises : une fois par le personnel de l'OQRE et une autre fois par un sous-traitant externe. Les résultats ont ensuite été comparés, puis les différences ont été corrigées, garantissant ainsi l'exactitude des résultats.

Une fois la notation terminée, des analyses statistiques et psychométriques ont été réalisées à l'aide de la théorie classique des scores et de la TRI afin de fournir des données sur la qualité technique de chacun des items de test ainsi que sur la fidélité globale des résultats des élèves.

Diverses statistiques de test ont été calculées, y compris le coefficient de fidélité alpha de Cronbach, les erreurs-types de mesure, les courbes caractéristiques de test, les fonctions informations de test, les statistiques relatives au fonctionnement différentiel de l'item ainsi que l'exactitude et l'uniformité de la classification, pour évaluer le degré de précision des résultats attribués aux tests de l'OQRE de 2008. Dans l'ensemble, les résultats de ces mesures indiquent que des niveaux de précision satisfaisants ont été obtenus. Les coefficients de fidélité variaient entre 0,78 et 0,89 pour les tests des cycles primaire et moyen, entre 0,81 et 0,85 pour le Test de mathématiques, 9^e année, et entre 0,86 et 0,88 pour le TPCL. L'exactitude de la classification, pour les élèves ayant un rendement égal ou supérieur à la norme provinciale ou, au TPCL, pour les élèves ayant réussi le test, dont le coefficient variait entre 0,86 et 0,92, indique qu'environ 90 % des élèves avaient été classifiés correctement.

Publication des résultats

Les résultats décrivent le rendement des élèves à des étapes déterminantes de leur éducation.

Pour les tests des cycles primaire et moyen et de 9^e année, une échelle à quatre niveaux est utilisée pour communiquer le rendement des élèves en lecture, en écriture et en mathématiques.

Pour le TPCL, deux niveaux de rendement sont communiqués : réussite et non réussite.

Tous les rapports sont conçus avec la plus grande attention en ce qui a trait à l'exactitude, à la clarté et à l'efficacité.

Les rapports comprennent :

- **des tableaux des résultats;**
- **des diagrammes des résultats;**
- **des renseignements directeurs pour l'interprétation des résultats.**

Les résultats des tests sont communiqués au niveau des élèves, des écoles, des conseils scolaires et de la province. Ils fournissent de l'information utile permettant de soutenir la planification de l'amélioration à tous ces niveaux.

Les niveaux de rendement utilisés pour rendre compte des résultats des tests des cycles primaire et moyen et de 9^e année proviennent du curriculum de l'Ontario, qui établit le niveau 3 comme la norme provinciale. Les niveaux 1 et 2 indiquent un rendement inférieur à la norme provinciale et le niveau 4 indique un niveau supérieur à cette norme.

Quant au TPCL, les élèves sont classés en deux catégories : réussite ou non réussite. Les élèves ayant réussi ont atteint le niveau de compétences linguistiques prévu à la fin de la 9^e année. Ils ont donc satisfait à l'exigence linguistique permettant d'obtenir leur diplôme. Les élèves n'ayant pas réussi reçoivent une rétroaction afin de les aider à travailler en vue d'atteindre la norme.

Les parents et les tutrices et tuteurs reçoivent les Rapports individuels de l'élève qui fournissent les résultats généraux de chaque élève.

Les rapports à l'intention des conseils scolaires et des écoles sont affichés sur le site Web public de l'OQRE à l'adresse suivante : www.oqre.on.ca. Si le nombre d'élèves pour une école ou un conseil scolaire donné est inférieur à 15, les résultats ne sont pas communiqués afin d'éviter la possibilité de pouvoir identifier les résultats individuels des élèves. Les fichiers de données et les rapports détaillés incluant les résultats qui ont été supprimés des rapports publics et les résultats des sous-groupes d'élèves (p. ex., élèves inscrits à un programme d'actualisation du français [ALF] ou de perfectionnement du français [PDF] et élèves ayant des besoins particuliers) sont affichés dans la section sécurisée du site Web de l'OQRE, qui est accessible au personnel des conseils scolaires et des écoles à l'aide d'un code d'identification et d'un mot de passe.

En outre, l'OQRE publie un rapport provincial sur chaque test, versions française et anglaise, à l'intention des intervenantes et intervenants en milieu scolaire et du grand public. Ces rapports sont accessibles sur le site Web public de l'OQRE.

Les rapports publiés à l'intention des écoles, des conseils scolaires et de la province fournissent des informations contextuelles; les résultats pour chaque matière pour les écoles, les conseils scolaires et la province; les résultats selon le sexe de l'élève, pour les élèves bénéficiant d'un soutien ou d'un programme d'ALF/de PDF et les élèves ayant des besoins particuliers; les résultats d'une année à l'autre; les forces et les points devant être améliorés relativement au curriculum; les résultats sur l'information des items pour chaque élève, école, conseil scolaire et la province ainsi que les résultats issus des questionnaires à l'intention des élèves, du personnel enseignant et des directions d'école. Les résultats provenant des questionnaires à l'intention du personnel enseignant et des directions d'école ne sont pas publiés au niveau des écoles.