

## **Guide d'utilisation des microdonnées**

### **Enquête sur les technologies de l'information et des communications dans les écoles**

**– 2003-2004 –**

Avril 2005

## Table des matières

1.0 Administration .....	3
2.0 Loi autorisant la collecte.....	3
3.0 Contexte .....	3
4.0 Objectifs.....	4
5.0 Contenu.....	4
6.0 Utilisations.....	5
7.0 Collecte des données.....	6
8.0 Traitement des données.....	7
8.1 Saisie des données.....	7
8.2 Vérification des données et détection des valeurs aberrantes.....	7
8.3 Codage des réponses.....	9
8.4 Création de variables dérivées .....	9
9.0 Techniques d'enquête .....	10
9.1 Population cible .....	10
9.1.1 <i>La base de sondage</i> .....	10
9.1.2 <i>Écoles des Premières nations</i> .....	11
9.2 Plan d'échantillonnage.....	12
9.3 Composition de la population cible .....	12
10.0 Non-réponse.....	13
11.0 Traitement de la non-réponse.....	14
11.1 Catégories de correction et pondérations initiales .....	15
11.2 Étalonnage.....	17
12.0 Évaluation de la qualité.....	18
12.1 Erreurs non dues à l'échantillonnage.....	18
12.2 Champ d'observation.....	19
12.3 Non-réponse.....	19
12.4 Saisie des données.....	24
12.5 Vérification des données et repérage des valeurs aberrantes.....	24
13.0 Lignes directrices pour l'analyse, la diffusion et les totalisations .....	25
13.1 Lignes directrices pour l'arrondissement.....	25
13.2 Lignes directrices pour la pondération.....	26
13.3 Estimations nominales ou quantitatives.....	28
13.3.1 <i>Estimations nominales</i> .....	28
13.3.2 <i>Estimations quantitatives</i> .....	29
13.4 Lignes directrices pour la diffusion du coefficient de variation .....	32
14.0 Calcul de la variance.....	35
14.1 Importance de la variance .....	35
14.2 Module de type Excel pour l'extraction de coefficients de variation .....	36
14.3 Calculer les limites de confiance à l'aide du coefficient de variation .....	39
14.4 Tests d'hypothèse.....	40
14.5 Coefficients de variation d'estimations quantitatives .....	41
Annexe A : Questionnaire.....	42
Annexe B : Caractéristiques d'une école .....	43

## 1.0 Administration

Parrainé par le programme *Rescol* d'Industrie Canada, en collaboration avec les principaux partenaires au Canada – gouvernements provinciaux et territoriaux, associations éducatives, commissions scolaires, écoles, enseignants et élèves – et menée par Statistique Canada, l'Enquête sur les technologies de l'information et des communications dans les écoles (ETICE) vise à mesurer l'accès aux TIC et leur intégration dans le milieu scolaire. Bibliothèque et Archives Canada a apporté son soutien à cette initiative.

Pour toute question au sujet de l'ensemble des données ou de son utilisation, veuillez communiquer avec :

### Statistique Canada

Services à la clientèle  
Centre de la statistique de l'éducation  
Téléphone : (613) 951-7608 ou composez sans frais 1 800 307-3382  
Télécopieur : (613) 951-9040  
Adresse électronique : [educationstats@statcan.ca](mailto:educationstats@statcan.ca)

## 2.0 Loi autorisant la collecte

L'Enquête sur les technologies de l'information et des communications dans les écoles (ETICE) a été menée en vertu de la *Loi sur la statistique*, chapitre S-19, Lois révisées du Canada, 1985. Les plans de collecte pour l'enquête sont conformes aux exigences de la circulaire du Conseil du Trésor 1986-19, Collecte de renseignements et recherche sur l'opinion publique fédérales, et sont enregistrés sous le matricule de collecte n° STC/ECT-165-75342.

## 3.0 Contexte

L'ETICE a été acheminée aux directeurs en vue de recueillir des données de base dignes de confiance sur la connectivité des écoles et sur divers aspects de l'accès aux TIC dans les écoles élémentaires et secondaires partout au Canada. L'enquête a été élaborée par le programme *Rescol* du gouvernement du Canada, en collaboration avec le Conseil consultatif national de Rescol et Statistique Canada.

En 2002, le Conference Board du Canada a élaboré un cadre analytique pour mesurer la connectivité et l'intégration des TIC dans les écoles. L'ETICE portait principalement sur deux éléments de ce cadre analytique, soit l'infrastructure et la portée. L'infrastructure

couvre les différentes composantes des TIC qui forment la base d'une école branchée; il s'agit, par exemple, du nombre d'ordinateurs et de leurs caractéristiques. La portée, quant à elle, renvoie au degré d'accès qu'ont les enseignants et les élèves à l'infrastructure des TIC.

## 4.0 Objectifs

Les données serviront à évaluer l'état actuel de l'infrastructure des TIC et son accessibilité, de même que certaines tendances quant à leur utilisation dans les écoles, dans le contexte canadien. Les résultats de cette enquête orienteront l'élaboration de politiques et programmes à venir et jetteront la base de nouvelles études sur l'utilisation des TIC dans l'apprentissage. L'information permettra aussi de mesurer l'ampleur et la qualité de l'infrastructure des TIC dans les écoles au Canada et de mettre en oeuvre des programmes et des politiques spécifiquement adaptés à ces écoles.

## 5.0 Contenu

La participation à cette enquête était volontaire. Les questionnaires ont été envoyés par la poste à tous les directeurs d'école en octobre-novembre 2003. Les écoles des Premières nations ont fait l'objet d'un envoi distinct en avril-mai 2004. Les répondants pouvaient renvoyer leur questionnaire par la poste ou y répondre en ligne. Le questionnaire est reproduit à l'annexe A.

### **Le questionnaire était divisé en 13 sections :**

- 1) *Renseignements sur l'école* : Renseignements sur le nombre d'enseignants en équivalents temps plein par sexe, le nombre d'élèves par sexe et par année, le nombre de locaux scolaires, le nombre d'employés en équivalents temps plein affectés à la bibliothèque de l'école, les dépenses annuelles pour la collection de la bibliothèque, les sources de financement des dépenses de la bibliothèque et l'emplacement de l'école.
- 2) *Matériel informatique* : Renseignements sur l'infrastructure matérielle des TIC disponible dans les écoles, comme le nombre d'ordinateurs disponibles à des fins éducatives, le niveau de rendement de ces ordinateurs et l'accès par les élèves à ces ordinateurs à l'extérieur des heures d'enseignement.
- 3) *Logiciels* : Renseignements sur les types de logiciels disponibles dans l'école et l'accès à ces logiciels par les élèves.
- 4) *Connexions Internet et intranet* : Renseignements sur la connectivité à l'Internet et à un intranet dans l'école.

- 5) *Internet et courrier électronique* : Renseignements sur l'utilisation de l'Internet et du courrier électronique dans l'école.
- 6) *Site Web de l'école* : Renseignements sur le site Web de l'école, le cas échéant.
- 7) *Vidéoconférence* : Renseignements sur la fréquence d'utilisation de la vidéoconférence dans l'école.
- 8) *Cours en ligne pour les élèves* : Renseignements sur l'utilisation de cours en ligne pour les élèves inscrits.
- 9) *Compétences et perfectionnement professionnel des enseignants* : Renseignements sur les compétences et le perfectionnement professionnel des enseignants dans les écoles.
- 10) *Soutien technique* : Renseignements sur le soutien technique pour les TIC dans les écoles.
- 11) *Plans et politiques des TIC* : Renseignements sur les plans et les politiques pour les TIC dans les écoles ou les conseils scolaires.
- 12) *Attitudes envers les TIC* : Renseignements sur l'opinion du directeur à propos de certains énoncés sur l'utilisation des TIC dans les écoles.
- 13) *Obstacles à l'utilisation des TIC* : « Principaux » obstacles qui entravent l'utilisation des TIC dans l'école, selon le directeur.

## 6.0 Utilisations

Les fichiers de microdonnées ont été mis à la disposition de l'arrondissement ou de la commission scolaire (selon le cas), du ministère de l'Éducation provincial/territorial, d'Industrie Canada et de Bibliothèque et Archives Canada.

Industrie Canada, par le truchement du programme *Rescol*, utilisera l'information pour évaluer l'état actuel de l'infrastructure des TIC et son accessibilité, de même que les tendances de son utilisation dans les écoles, dans le contexte canadien. L'information de cette enquête orientera l'élaboration de politiques et programmes à venir et jettera la base de nouvelles études sur l'utilisation des TIC dans l'apprentissage. Bibliothèque et Archives Canada s'intéresse aux données dans le but de mieux comprendre l'état actuel de l'infrastructure de toutes les bibliothèques scolaires et de mettre en oeuvre de nouvelles politiques. L'information qui sera fournie à l'arrondissement ou commission scolaire et au ministère de l'Éducation provincial ou territorial leur permettra de mesurer l'ampleur et la qualité de l'infrastructure des TIC dans l'école et de concevoir des

programmes et des politiques spécifiquement adaptés aux écoles de la province ou territoire ou de l'arrondissement ou commission scolaire.

En plus de la publication d'études analytiques liées à la présence des TIC dans les écoles, l'information permettra aux praticiens de l'éducation et aux décideurs de mieux cerner les secteurs où une intervention et des mesures de soutien supplémentaires s'imposent.

## 7.0 Collecte des données

La collecte des données a eu lieu d'octobre 2003 à février 2004. Les répondants de cette enquête étaient les directeurs qui ont fourni à la fois les données de même que leurs perceptions quant aux TIC. Dans certains cas, cependant, les directeurs peuvent avoir répondu après avoir consulté d'autres intervenants ou discuté avec eux.

Les répondants pouvaient utiliser le questionnaire papier ou la collecte électronique de données (CED). On a attribué à chaque répondant un numéro d'identification et un mot de passe uniques, imprimés sur le questionnaire avec le nom, l'adresse et le numéro de téléphone de l'école. On demandait aux répondants soit de renvoyer le questionnaire rempli par la poste dans l'enveloppe fournie, soit de fournir les données en ligne.

Même si la participation à l'enquête était volontaire, une télécopie de rappel a été envoyée aux répondants, suivie d'appels téléphoniques pour les inciter à participer. Lorsque l'appel de l'intervieweur n'était pas opportun, on prenait rendez-vous pour rappeler à un autre moment.

Si les répondants n'avaient absolument pas le temps de remplir le questionnaire, on leur offrait alors la possibilité de répondre à un ensemble de questions critiques (questions 2, 3, 6, 15, 16 et 49) et de consentir à la divulgation des renseignements au sujet de leur école.

Question 2 – Nombre d'enseignants en équivalents temps plein par sexe

Question 3 – Nombre d'élèves par sexe et année

Question 6 – Nombre de locaux scolaires par type

Question 15 – Nombre d'ordinateurs de bureau par type

Question 16 – Nombre d'ordinateurs portatifs et de blocs-notes par type

Question 49 – Obstacles perçus à l'utilisation des TIC

Chaque questionnaire reçu était chargé dans Blaise (un système de collecte) et des vérifications prédéfinies relevaient les données manquantes aux questions critiques. Un intervieweur appelait alors le répondant afin d'obtenir l'information manquante. Après avoir parlé au répondant, l'intervieweur pouvait attribuer à chacune des questions le code *Refus* ou *Ne sait pas*. Les vérifications liées aux questions 2, 3, 6, 15, 16 et la question du consentement étaient *impératives*, c.-à-d. qu'il devait y avoir un suivi auprès du

répondant pour qu'il explique l'écart ou fournisse l'information manquante. Le nombre de suivis auprès des écoles pour corriger les valeurs déclarées par les répondants a été fortement limité par les échéances imposées après la période de collecte.

## 8.0 Traitement des données

Ce chapitre présente un bref aperçu des étapes du traitement qui ont mené à la production du fichier de microdonnées.

### 8.1 Saisie des données

Les réponses aux questions de l'enquête ont été saisies par l'une ou l'autre de deux méthodes : la collecte électronique de données (CED) et la reconnaissance intelligente des caractères (RIC). Les répondants qui utilisaient la CED entraient directement leurs données dans l'application qu'ils retournaient ensuite à Statistique Canada par un site ftp (protocole de transfert de fichier) sécurisé. S'ils répondaient sur le questionnaire papier, les données étaient alors saisies par RIC. La technologie de la RIC allie l'*entrée des données automatisée* (basée sur la reconnaissance optique des caractères, coches et images) à une *saisie manuelle* supplémentaire sur clavier par des préposés qui « *saisissent de l'image* » une partie des informations de l'enquête au moyen d'un affichage frontal.

Pour garantir la qualité des données saisies par la RIC, toutes les zones à remplir faisaient l'objet d'une double saisie au clavier pour en assurer l'exactitude puis on recourait à une procédure de contrôle qualitatif qui consistait, pour chaque lot de questionnaires saisis, à prélever 20 % des questionnaires pour en comparer l'image aux données réelles.

Des vérifications d'intervalle ont été combinées à la saisie par la CED et la RIC. Si les renseignements introduits se situaient à l'extérieur de l'intervalle (trop grands ou trop petits) des valeur attendues, ou étaient incohérents, les données étaient alors vérifiées et modifiées, ou refusées.

### 8.2 Vérification des données et détection des valeurs aberrantes

Parmi les étapes du traitement, la vérification des données et la détection des valeurs aberrantes cernent les valeurs erronées ou incohérentes dans les données d'enquête et les modifient.

Le premier type d'erreur traité porte sur l'enchaînement des questions. Des renseignements contradictoires pourraient parfois indiquer que des répondants ont répondu à des questions d'une section qui, de fait, ne s'appliquaient pas à eux. Dans ces cas, on a éliminé les données superflues. Voici un exemple de ce type d'erreur : un

répondant indique aux questions 15 et 16 que les ordinateurs sont connectés à l'Internet tout en répondant à la question 23 qui demande pourquoi l'école n'a pas de connexion Internet.

Le deuxième type d'erreur traité consistait à vérifier les enregistrements de l'enquête selon des règles préétablies afin de déceler les incohérences logiques dans les données d'enquête. Dans ces cas, une procédure éliminait un minimum de données grâce à des priorités qui supprimaient les valeurs le plus susceptibles d'être erronées. Voici un exemple : à la question 8, le nombre d'enseignants affectés à la bibliothèque est plus élevé que le nombre total d'enseignants dans l'école indiqué à la question 2. Dans ce cas, la réponse à la question 2 serait plus susceptible d'être bonne et les cellules fautives de la question 8 seraient remplacées par un code « Non déclaré ».

La troisième composante du processus de vérification visait le traitement des réponses multiples aux questions du type de l'échelle de Likert (questions 13, 21 et 47, par exemple). À ces questions, des répondants cochaient plus d'une case (p. ex., « Quelques-uns » et « Plusieurs » à la question 18). Pour chacun de ces enregistrements non valides, on choisissait au hasard une des catégories indiquées. La procédure établie tenait compte de la fréquence des catégories choisies dans les enregistrements valides (une seule case cochée). Par exemple, si les deux choix « Quelques-uns » et « Plusieurs » étaient cochés dans un enregistrement non valide pour un élément à la question 18, et que ces deux choix revenaient 25 fois et 75 fois respectivement parmi les enregistrements valides, la valeur « Quelques-uns » avait une chance sur quatre d'être attribuée à l'enregistrement non valide et la valeur « Plusieurs », trois chances sur quatre.

Le quatrième type d'erreur consistait à attribuer le code « Non déclaré » lorsque les répondants n'avaient fourni aucune information aux questions qui auraient dû être répondues.

Enfin, le dernier type d'erreur portait sur le repérage des valeurs aberrantes. Des méthodes statistiques, permettant de trouver des écarts dans la répartition ou des valeurs extrêmement éloignées de la médiane, ont servi à vérifier la distribution des variables numériques et à cerner les valeurs qui étaient suffisamment éloignées de la médiane pour justifier une vérification. On a appliqué ces méthodes non seulement aux variables saisies dans l'enquête mais aussi aux ratios découlant des variables de l'enquête (p. ex., les ratios élèves-ordinateurs, élèves-enseignants) et aux ratios des variables de l'enquête à celles de la base de sondage (p. ex., le nombre d'élèves relevés à partir de l'enquête au nombre d'élèves présents dans la base de sondage). Les valeurs aberrantes extrêmes ont été repérées et comparées aux images des questionnaires correspondants pour vérifier si les données avaient été saisies correctement. On a contacté les répondants dont les réponses contenaient des valeurs aberrantes extrêmes qui semblaient avoir été saisies correctement. En bout de ligne, on a compté environ 50 enregistrements dont les valeurs aberrantes ont été confirmées ou modifiées. Le nombre de suivis a été fortement limité par les échéances imposées après la période de collecte.



### 8.3 Codage des réponses « Autre » – « Veuillez préciser »

Quelques données élémentaires du questionnaire ont fait l'objet de questions ouvertes, comme « Autre – Veuillez préciser ». Ce type de réponse survient lorsqu'une question comporte comme réponse une liste de catégories possibles de même qu'une possibilité de réponse autre que celles déjà énumérées dans la liste. Dans cette situation, le texte a été saisi puis vérifié manuellement. Si la réponse figurait dans l'une des catégories existantes, elle était recodée dans la catégorie appropriée. Les autres sont demeurées de véritables réponses « Autre – Veuillez préciser ».

### 8.4 Création de variables dérivées

On a dérivé un certain nombre de variables du fichier de microdonnées en regroupant des éléments du questionnaire ou en utilisant des renseignements de la base de sondage afin de faciliter l'analyse des données. On peut ainsi utiliser une variable ou une combinaison de variables. La liste des variables dérivées de l'ETICE figure ci-après. Vous trouverez à l'annexe B une description plus détaillée des variables dérivées liées aux caractéristiques de l'école (le niveau d'enseignement de l'école, la taille de l'établissement, l'emplacement de l'école et aussi le financement de l'école.)

Variable	Description
RuralDChosen	Emplacement de l'école (région urbaine, rurale)
TotalMale_D	Nombre d'élèves de sexe masculin dans toutes les années
TotalFemale_D	Nombre d'élèves de sexe féminin dans toutes les années
Total_D	Nombre d'élèves des deux sexes dans toutes les années
Typeschd	Financement de l'école (public, privé, mixte)
MINGRADE	Niveau d'année minimal de l'école
MAXGRADE	Niveau d'année maximal de l'école
GRADLVD	Niveau d'enseignement de l'école (élémentaire, secondaire, mixte)
Schized	Taille de l'école (petite, moyenne, grande)
Numcompd_Q15	Nombre d'ordinateurs de bureau
Numcompd_Q16	Nombre d'ordinateurs portatifs et de blocs-notes
Numcompd	Nombre d'ordinateurs
Intcond	Nombre d'ordinateurs branchés à Internet
Nintcond	Nombre d'ordinateurs non branchés à Internet
Lowspdd	Nombre d'ordinateurs avec processeur à faible vitesse
Medspdd	Nombre d'ordinateurs avec processeur à vitesse moyenne
Highspdd	Nombre d'ordinateurs avec processeur d'une autre vitesse
Otherspdd	Nombre d'ordinateurs avec processeur à vitesse autre
Locclasd	Nombre d'ordinateurs dans les salles de cours/classes mobiles
Loclabd	Nombre d'ordinateurs dans les laboratoires informatiques
Loclibd	Nombre d'ordinateurs dans les bibliothèques
Locothd	Nombre d'ordinateurs dans d'autres endroits
Stcomrd	Ratio élèves-ordinateurs

---

Lowispdd	Nombre d'ordinateurs branchés à Internet avec processeur à faible vitesse
Medispdd	Nombre d'ordinateurs branchés à Internet avec processeur à vitesse moyenne
Highispdd	Nombre d'ordinateurs branchés à Internet avec processeur à haute vitesse
Otherispdd	Nombre d'ordinateurs branchés à Internet avec processeur d'une autre vitesse
Lociclad	Nombre d'ordinateurs branchés à Internet dans les salles de cours/classes mobiles
Locilabd	Nombre d'ordinateurs branchés à Internet dans les laboratoires informatiques
Locilibd	Nombre d'ordinateurs branchés à Internet dans les bibliothèques
Lociothd	Nombre d'ordinateurs branchés à Internet dans d'autres endroits
Sticomrd	Ratio élèves-ordinateurs branchés à Internet
Desklcond	Nombre d'ordinateurs de bureau branchés à Internet
Laplcond	Nombre d'ordinateurs portatifs et blocs-notes branchés à Internet
Totlan	Écoles avec système d'exploitation pour un réseau local
Old_os	Pourcentage d'ordinateurs fonctionnant sous un système d'exploitation plus ancien
New_os	Pourcentage des ordinateurs fonctionnant sous un système d'exploitation récent
ProvCode	Code de la province ou du territoire

---

## 9.0 Techniques d'enquête

Ce bref chapitre décrit les écoles considérées comme étant dans le champ de l'enquête et le nombre final d'unités dans la base de sondage pour chaque province et territoire.

### 9.1 Population cible

La population cible de l'ETICE englobe toutes les écoles primaires et secondaires au Canada, sauf les écoles d'éducation permanente ou de jour pour adultes, les écoles professionnelles ou de métiers, les écoles de langue et d'éducation culturelle, les écoles à domicile, les centres d'éducation communautaire ainsi que les centres de services sociaux. L'enquête inclut toutes les écoles situées dans toutes les provinces et tous les territoires, y compris celles situées dans les communautés autochtones. On considérerait dans le champ de l'ETICE seulement les écoles ouvertes durant l'année scolaire 2002-2003.

#### *9.1.1 La base de sondage*

On peut avoir accès à la population cible grâce à la base de sondage. La base de sondage utilisée pour l'ETICE est le fichier des établissements pour 2002 : une base de données

administratives mise à jour par le Centre de la statistique de l'éducation (CES) de Statistique Canada. Le fichier des établissements renferme un ensemble de renseignements sur chaque école, notamment la personne-ressource, le niveau d'année minimal et maximal, le financement de l'école et le nombre d'élèves dans l'école.

Le CES rassemble ses informations à partir des sources les plus à jour qui déclarent des données annuelles sur les écoles élémentaires et secondaires au Canada, afin d'actualiser le fichier des établissements de l'année précédente. Des demandes sont envoyées aux ministères provinciaux et/ou aux directeurs d'école pour mettre à jour l'information de chaque école dans le fichier. Le fichier des établissements utilisé pour créer la base de sondage de l'ETICE contenait toutes les mises à jour reçues jusqu'à l'automne de 2003 à partir des demandes envoyées à l'automne de 2002.

Des recherches supplémentaires ont été menées à l'automne de 2003 afin de disposer des informations les plus récentes possibles avant la tenue de l'enquête. Ces recherches visaient principalement à retirer de la liste le plus d'enregistrements possibles qui étaient hors du champ de l'enquête ou présents en double. Par exemple, des enregistrements indiquaient qu'une école n'avait pas d'élèves ou que des adultes y recevaient un enseignement, même si l'on était en présence d'un type d'école valide. Après examen, des enregistrements ont été supprimés de la base de sondage parce qu'ils étaient hors du champ de l'enquête.

D'autres enregistrements de la base de sondage ne contenaient pas des données analytiques clés. On a cherché ces valeurs manquantes soit en ligne, soit dans le répertoire des écoles canadiennes de Scott. Pendant cette recherche, il est devenu évident qu'il subsistait toujours des enregistrements hors du champ de l'enquête. Quatre autres enregistrements comportaient des annotations indiquant qu'ils étaient hors du champ de l'enquête – cette information a été confirmée et ces écoles furent retirées de la base de sondage.

Afin de repérer les doubles qui devaient être retirés de la liste, la base de sondage a fait l'objet d'un tri à l'aide de variables qui, en combinaison, auraient dû être des identificateurs uniques dans la base de sondage. Ces variables comprenaient le nom, l'adresse, le code postal et le numéro de téléphone de l'école.

Après avoir rassemblé et analysé les données, on a décidé que les écoles virtuelles et les écoles à distance (comme les écoles à domicile) présentaient des caractéristiques bien différentes de celles des autres types d'école contactés (sur le plan du ratio élèves-ordinateurs et de l'utilisation d'ordinateurs en ligne, entre autres caractéristiques). Voilà pourquoi elles ont été retirées du champ de l'enquête.

### *9.1.2 Écoles des Premières nations*

Pour préparer l'envoi des questionnaires à toutes les écoles élémentaires et secondaires du Canada, on a demandé l'aide de diverses administrations. Nous avons contacté les

ministères de l'Éducation provinciaux en ce qui a trait à la majorité des écoles élémentaires et secondaires du Canada. Toutefois, afin de contacter les écoles des Premières nations au Canada, une permission spéciale devait être accordée par les différents organismes de gestion régionaux des Premières nations (RMO). Comme cela nécessitait plus de temps que de contacter les ministères provinciaux, on a exclu de l'envoi initial les écoles des Premières nations figurant dans la base de sondage.

Grâce à une liste des écoles des Premières nations au Canada qui avait été fournie à Statistique Canada par le programme *Rescol* des Premières nations à Industrie Canada et grâce aussi à la zone « financement » du fichier des établissements, nous avons pu retirer temporairement 400 écoles des Premières nations de la base de sondage. Les RMO ont par la suite accordé la permission de contacter les écoles des Premières nations et les questionnaires de l'ETICE ont été envoyés en avril 2004.

## 9.2 Plan d'échantillonnage

Comme nous l'avons mentionné, toutes les écoles élémentaires et secondaires du Canada qui étaient dans le champ de l'enquête, tel que décrit précédemment dans ce chapitre, ont été contactées pour participer.

## 9.3 Composition de la population cible

Les tableaux 1 et 2 donnent un aperçu de la répartition des écoles contactées en octobre 2003 dans la vague principale de l'enquête (qui exclut les écoles des Premières nations), par province et territoire et selon les cinq principales variables analytiques (niveau d'enseignement, taille, emplacement, financement et langue d'enseignement de l'école).

Tableau 1

### **Nombre d'écoles contactées pour l'ETICE, par province et territoire**

<b>Province/territoire</b>	<b>Population</b>
Terre-Neuve	331
Île-du-Prince-Édouard	72
Nouvelle-Écosse	475
Nouveau-Brunswick	366
Québec	3 013
Ontario	5 634
Manitoba	820
Saskatchewan	808
Alberta	1 880
Colombie-Britannique	2 020
Yukon	29
Territoires du Nord-Ouest	52
Nunavut	41

---

<b>Total</b>	15 541
--------------	--------

---

Tableau 2

**Nombre d'écoles contactées pour l'ETICE, par catégorie définie selon les variables analytiques clés**

<b>Domaine</b>		<b>Population</b>
<b>Niveau d'enseignement</b>	Élémentaire	10 121
	Secondaire	3 404
	Mixte	2 016
<b>Taille</b>	Petite	4 970
	Moyenne	5 246
	Grande	5 325
<b>Emplacement</b>	Urbain	11 455
	Rural	4 086
<b>Financement</b>	Public	13 704
	Privé	1 666
	Mixte	171
<b>Langue</b>	Anglaise	12 225
	Française	3 316

## 10.0 Non-réponse

Les taux de réponse d'une enquête mesurent l'efficacité du processus de collecte et sont aussi de bons indicateurs de la qualité des estimations produites. Peut-être plus que d'autres enquêtes, l'ETICE souffre de plusieurs niveaux de non-réponse. Ce chapitre présente un sommaire qui distinguera trois types de non-réponse : non-réponse totale et partielle et non-réponse à une question.

### **Non-réponse totale**

Pour l'unité échantillonnée, l'*ensemble minimal de questions critiques* n'a pas été recueilli. Cet « ensemble minimal de questions critiques » fait référence à l'ensemble des questions critiques définies au chapitre 7.0, à l'exception des questions 2 (le nombre d'enseignants) et 6 (le nombre de locaux scolaires) dont on a jugé qu'elles étaient moins essentielles au plan analytique que le compte des effectifs ou le consentement à divulguer l'information. De même, une seule des questions 15 ou 16 nécessitait une réponse car on peut concevoir facilement qu'une école dispose d'ordinateurs de bureau mais non d'ordinateurs portatifs (ou vice-versa). On a utilisé des méthodes de pondération pour compenser la non-réponse totale; plus de détails figurent à ce sujet au chapitre 11.0 (veuillez vous référer à l'étape 1, Corrections de pondération).

### **Non-réponse partielle**

On a recueilli uniquement les données liées à l'ensemble minimal des variables critiques définies pour la non-réponse totale. Des méthodes de pondération ont été utilisées pour compenser la non-réponse partielle. On développe ce sujet au chapitre 11.0.

### **Non-réponse à une question**

On pourrait associer la non-réponse à une question à n'importe quel questionnaire, à l'exception de ceux qui composent l'ensemble de la non-réponse totale. La non-réponse à une question est une absence limitée d'information pour un petit nombre de variables, cependant qu'on a recueilli des données pour toutes les autres variables dans le questionnaire (y compris l'ensemble minimal de questions critiques). On corrige habituellement la non-réponse à une question par des techniques d'imputation, mais dans le cas de l'ETICE, les taux de non-réponse à des questions étaient suffisamment bas pour la plupart des variables pour éviter ce processus.

Des 15 541 écoles du tableau 1, 6 676 ont fourni une information utilisable aux fins de l'enquête (c.-à-d. que ces questionnaires ne faisaient pas partie de l'ensemble de la non-réponse totale), pour un taux de réponse de 43 %. Celles-ci forment un sous-ensemble de 7 311 questionnaires renvoyés (un taux de retour de 47 %). Un sous-ensemble (6 103 questionnaires) de ces questionnaires utilisables présentaient des réponses au-delà des seules questions critiques et, par conséquent, aucune non-réponse partielle (pour un taux de réponse complet de 39 %). Pour plus d'information sur ces questions de non-réponse, veuillez vous référer au chapitre 11.0 et à la section 12.3.

## **11.0 Traitement de la non-réponse**

La non-réponse est la principale source d'erreur dans une enquête comme l'ETICE. Les erreurs de cette nature n'étant généralement pas aléatoires, il importe de la réduire et aussi d'adopter une stratégie de correction appropriée qui tient compte de la non-réponse systématique. À la lumière de l'analyse des taux de réponse, et compte tenu des contraintes opérationnelles, on a examiné les schémas de non-réponse et la stratégie de correction en deux étapes :

Étape 1 : Examen des questions critiques seulement pour l'ensemble des répondants (c.-à-d. y compris les questionnaires présentant une non-réponse partielle).

Étape 2 : Examen de toutes les questions uniquement pour les écoles ayant répondu au-delà des questions critiques (c.-à-d. en excluant les questionnaires présentant une non-réponse partielle).

À chaque étape, la pondération attribuée à chaque école représente le nombre des autres écoles de la population ayant des caractéristiques similaires. Les pondérations ont été calculées en deux temps, comme le décrit la section 11.1.

On a déterminé que ces ensembles de données comportaient de faibles taux de non-réponse à la majorité des questions critiques (étape 1) et aux autres questions restantes (étape 2). On a donc adopté une méthodologie de pondération fondée sur des renseignements complémentaires importants consignés dans la base de sondage (c.-à-d. province ou territoire, niveau d'enseignement, taille, emplacement, financement et langue d'enseignement de l'école) afin de corriger la non-réponse partielle et totale).

Malheureusement, quelques enregistrements de la base de sondage ne comportaient pas les informations analytiques clés requises pour faire cette correction. Dans ces cas-là, les données manquantes ont été imputées d'après l'information disponible dans des enregistrements complets similaires. Voilà pourquoi nous avons imputé le nombre d'élèves et le niveau d'enseignement respectivement à 3,1 % et 0,3 % des enregistrements de la base de sondage.

### 11.1 Catégories de correction et pondérations initiales

On a effectué une analyse exploratoire, selon le modèle de régression logistique, pour étudier le rapport entre l'information complémentaire clé disponible (province ou territoire, niveau d'enseignement, taille, emplacement, financement et langue d'enseignement de l'école) dans la base de sondage et la probabilité de réponse des écoles. C'est ainsi que les écoles ont été groupées en plusieurs catégories de correction qui nous permettaient de présumer que chaque unité d'une catégorie avait la même probabilité de réponse. On a défini comme suit la pondération initiale attribuée à toutes les écoles qui appartenaient à une catégorie de correction :

#### **Catégories de correction de l'étape 1 :**

$$W(\text{étape 1}) = \frac{\text{Nombre d'écoles dans cette catégorie, tirées de la base de sondage}}{\text{Nombre d'écoles dans cette catégorie qui ont répondu à l'enquête}}$$

#### **Catégories de correction de l'étape 2 :**

$$W(\text{étape 2}) = W(\text{étape 1}) \times \frac{\text{Nombre d'écoles dans cette catégorie qui ont répondu à l'enquête}}{\text{Nombre d'écoles dans cette catégorie qui ont répondu au-delà des questions critiques}}$$

Le tableau 3 énumère les groupes de correction pour la non-réponse de l'étape 1 et les taux de non-réponse totale correspondants. On calcule les taux comme une proportion des valeurs attendues dérivées de la base de sondage de l'ETICE.

Tableau 3

**Non-réponse totale en fonction des groupes de correction pour la non-réponse de l'étape 1**

Groupe	Base de sondage	Non-réponse totale		Groupe	Base de sondage	Non-réponse totale	
		Nombre	Taux			Nombre	Taux
Terre-Neuve	331	133	40,2 %	Sask., secondaire moyenne/grande	66	28	42,4 %
Î.-P.-É.	72	20	27,8 %	Sask., secondaire petite	70	27	38,6 %
Nouvelle-Écosse	475	151	31,8 %	Alb., élémentaire grande	233	81	34,8 %
Nouveau-Brunswick	366	78	21,3 %	Alb., élémentaire moyenne	304	138	45,4 %
Québec, langue anglaise	353	198	56,1 %	Alb., élémentaire petite	254	126	49,6 %
Québec, langue française	2 660	1 702	64,0 %	Alb., mixte grande	239	125	52,3 %
Ont., langue anglaise, urbaine	4 317	2 881	66,7 %	Alb., mixte moyenne	176	88	50,0 %
Ont., langue anglaise, rurale	928	560	60,3 %	Alb., mixte petite	205	136	66,3 %
Ont., langue française, urbaine	316	178	56,3 %	Alb., secondaire grande	93	38	40,9 %
Ont., langue française, rurale	73	39	53,4 %	Alb., secondaire moyenne	188	91	48,4 %
Man., public ou mixte	734	323	44,0 %	Alb., secondaire petite	188	112	59,6 %
Man., privé	86	53	61,6 %	C.-B., public	1 682	1 022	60,8 %
Sask., élémentaire grande	67	43	64,2 %	C.-B., privé	260	151	58,1 %
Sask., élémentaire moyenne	154	60	39,0 %	C.-B., financement mixte	78	35	44,9 %
Sask., élémentaire petite	153	70	45,8 %	Yukon	29	5	17,2 %
Sask., mixte, grande	65	24	36,9 %	Territoires du N.-O.	52	15	28,8 %
Sask., mixte, moyenne	135	53	39,3 %	Nunavut	41	18	43,9 %
Sask., mixte, petite	98	63	64,3 %				

Les groupes de correction pour la non-réponse créés à l'étape 2 et les taux de non-réponse partielle correspondants sont présentés au tableau 4. Les taux représentent la proportion de l'ensemble de tous les répondants.



Tableau 4

**Non-réponse partielle en fonction des groupes de correction pour la non-réponse de l'étape 2**

Groupe	Répondants	Non-réponse partielle		Groupe	Répondants	Non-réponse partielle	
		Nombre	Taux conditionnel			Nombre	Taux conditionnel
Terre-Neuve	198	11	5,6 %	Ont., secondaire	412	25	6,1 %
Î.-P.-É.	52	4	7,7 %	Manitoba	444	53	11,9 %
Nouvelle-Écosse	324	7	2,2 %	Saskatchewan	440	45	10,2 %
Nouveau-Brunswick	288	9	3,1 %	Alberta	945	75	7,9 %
QC, élémentaire	811	92	11,3 %	C.-B., urbaine	648	89	13,7 %
QC, niveau mixte	60	2	3,3 %	C.-B., rurale	164	12	7,3 %
QC, secondaire	242	11	4,5 %	Yukon	24	0	0,0 %
Ont., élémentaire	1 474	128	8,7 %	Territoires N.-O.	37	2	5,4 %
Ont., niveau mixte	90	2	2,2 %	Nunavut	23	6	26,1 %

La non-réponse partielle et totale fluctue principalement par province/territoire (voir le tableau 5 à la section 12.3 pour les taux de non-réponse provinciaux et nationaux) et c'est pourquoi les premières catégories créées correspondent chacune à une province ou à un territoire. À l'intérieur de quelques provinces/territoires, d'autres facteurs ont pu jouer sur les taux de réponse. Par exemple, au Québec, l'analyse de régression logistique indiquait que la réponse d'une école dépendait vraisemblablement du fait qu'elle était de langue française ou de langue anglaise (voir le tableau 3). C'est pourquoi le groupe Québec a été divisé en deux catégories : Québec/langue anglaise et Québec/langue française. Les autres catégories infraprovinciales ont été créées de la même façon.

## 11.2 Étalonnage

Les techniques d'estimation par étalonnage sont très utilisées dans les enquêtes sociales. Elles produisent des estimations des totaux pour les variables clés qui correspondent aux chiffres de population connus. Cette approche permet aussi d'améliorer la qualité des estimations de l'enquête s'il y a un rapport entre les variables clés utilisées dans l'étalonnage et les variables utilisées pour l'estimation.

Pour calculer les pondérations finales de l'étape 1, on a ajusté les pondérations initiales définies dans la section 11.1 afin que les totaux estimatifs pour les variables clés

disponibles dans la base de sondage équivalent aux totaux correspondants de la base de sondage. Pour l'étape 2, on a ajusté les pondérations initiales afin que les totaux estimatifs associés aux écoles qui avaient répondu au-delà des questions critiques (c.-à-d. les répondants de l'étape 2) concordent avec les estimations correspondantes dérivées au moyen des pondérations finales de l'étape 1 pour tous les répondants (c.-à-d. les répondants de l'étape 1). Pour les deux étapes, on a réduit les pondérations initiales en fonction du carré de l'écart entre les pondérations finales et les pondérations initiales (une technique dite d'estimation par régression généralisée).

Lorsque c'était possible, on a utilisé toutes les catégories des variables clés pour ajuster les pondérations (par exemple, pour les types de financement, on utiliserait idéalement le nombre estimatif d'écoles à financement public, privé et mixte pour étalonner les pondérations). Toutefois pour quelques provinces ou territoires, le petit nombre de répondants disponibles nous a obligés à regrouper quelques-unes des catégories (p. ex., en Ontario, on utilisé le nombre estimatif d'écoles à financement public **ou** mixte pour étalonner les pondérations).

## 12.0 Évaluation de la qualité

Ce chapitre permet à l'utilisateur de prendre connaissance des divers éléments qui influencent la qualité des données de l'enquête. Dans une enquête par sondage ordinaire, on distingue deux principaux types d'erreur : les erreurs d'échantillonnage et les erreurs non dues à l'échantillonnage. On appelle erreur d'échantillonnage la différence entre une estimation dérivée d'un échantillon et l'estimation qui aurait été obtenue à partir d'un recensement qui aurait utilisé les mêmes procédés de collecte. L'ETICE étant un recensement, aucune possibilité d'erreur d'échantillonnage n'existait. On appelle erreurs non dues à l'échantillonnage tous les autres types d'erreur, y compris les problèmes liés au champ d'observation de l'enquête, la non-réponse et les erreurs de traitement, qui sont toutes examinées dans la section ci-après.

### 12.1 Erreurs non dues à l'échantillonnage

Il y a de nombreuses sources d'erreurs non dues à l'échantillonnage et elles peuvent se produire à presque toutes les étapes d'une enquête. Par exemple : des employés responsables de la collecte comprennent mal les instructions; des répondants ne comprennent pas des questions; ou des réponses sont mal inscrites sur le questionnaire. Des erreurs peuvent être introduites à tout moment lors de la manipulation des données, y compris lors du dépouillement et de la totalisation des données.

Pour l'ETICE, on a adopté des mesures d'assurance de la qualité afin de réduire le plus possible la présence d'erreurs non dues à l'échantillonnage, par exemple : contacts initiaux avec les ministères provinciaux et territoriaux, les commissions scolaires, les organismes pertinents associés à la population scolaire visée et, il va de soi, les directeurs

d'école; le questionnaire a été mis à l'essai; des activités de suivi ont été mises en place après la collecte de données, y compris une formation adéquate du personnel sur le terrain; on a effectué des contrôles de validation pour vérifier la cohérence des données. Pour plus de renseignements sur les procédures de collecte et de traitement, veuillez vous référer aux chapitres 7.0 et 8.0. Quelques-unes de ces mesures indiquent l'étendue des erreurs non dues à l'échantillonnage associées à l'enquête et elles sont décrites ci-après.

## 12.2 Champ d'observation

On a évalué la qualité de la base de sondage en examinant le sous-dénombrement et le surdénombrement, de même que les enregistrements en double. Des difficultés sont survenues lors de l'évaluation du sous-dénombrement de la base de sondage, car on a découvert que la plupart des autres listes auxquelles on pouvait comparer le fichier des établissements étaient soit étroitement liées aux fichiers administratifs utilisés pour créer la base de sondage de l'ETICE, soit apparentées à une version antérieure de ces fichiers, soit dépourvues d'un indicateur commun. Toutefois, en raison du processus dynamique des suivis annuels auprès des écoles et de la comparaison annuelle opérée avec les chiffres globaux publiés par les ministères provinciaux, on s'entend généralement pour dire que le sous-dénombrement est assez faible.

Les procédés expliqués à la section 9.1.1 et utilisés pour repérer les enregistrements en double ou hors du champ de l'enquête ont permis de découvrir sept enregistrement en double et 24 enregistrements hors du champ de l'enquête; ces enregistrements ont été retirés de la base de sondage sans plus d'explications. Vu l'étendue de ces procédés et le nombre relativement restreint d'enregistrements nécessitant un retrait, on estime que la proportion des enregistrements en double et hors du champ de l'enquête qui sont passés inaperçus dans la base de sondage est négligeable.

Afin de déterminer la qualité du nombre d'élèves imputés dans la base de sondage (voir le chapitre 11.0), on a comparé les données imputées aux chiffres des inscriptions publiés par Statistique Canada pour l'année scolaire 1999-2000. Il en ressort que la qualité de l'imputation semble acceptable et la méthodologie, valable. En tenant compte de cela et de la proportion minimale de tels cas imputés (3,1 %), nous pouvons dire que l'imputation devrait avoir un impact négligeable sur les estimations de l'enquête.

## 12.3 Non-réponse

Si l'on peut dire sans se tromper que la non-réponse partielle et totale (définie au chapitre 10.0) se produit totalement de façon aléatoire, on peut alors l'ignorer. Toutefois, c'est rarement le cas; habituellement, le groupe de répondants et le groupe de non-répondants présentent des caractéristiques différentes et l'ignorance de ce fait peut produire un biais dans les estimations de l'enquête. Comme nous l'avons décrit au chapitre 11.0, les corrections de pondération ont servi à réduire le risque d'introduction d'un biais lié à la non-réponse dans les estimations de l'ETICE.

Le tableau 5 illustre la répartition des chiffres et des taux de non-réponse par province et territoire pour chaque étape, mettant en évidence les 8 865 écoles de la base de sondage dont les questionnaires n'étaient pas utilisables (non-réponse totale) et les 573 écoles qui ont répondu seulement aux questions critiques (non-réponse partielle). Le nombre d'écoles dans la colonne « base de sondage » sont les mêmes 15 541 écoles dans le champ de l'enquête que l'on trouve au tableau 1 (en excluant les écoles des Premières nations). Comme nous l'avons décrit au chapitre 10.0, les écoles du groupe de la non-réponse totale sont celles qui ont présenté un questionnaire non utilisable, car elles n'avaient pas fourni l'ensemble minimal des réponses aux questions critiques. Tous les taux sont calculés en proportion des chiffres attendus dérivés de la base de sondage de l'ETICE.

Tableau 5

**Non-réponse partielle et totale par province et territoire**

Province et territoire	Base de sondage	Non-réponse totale		Non-réponse partielle		Non-réponse totale et partielle	
		Nombre	Taux	Nombre	Taux	Nombre	Taux
T.-N.	331	133	40,2 %	11	3,3 %	144	43,5 %
Î.-P.-É.	72	20	27,8 %	4	5,6 %	24	33,3 %
N.-É.	475	151	31,8 %	7	1,5 %	158	33,3 %
N.-B.	366	78	21,3 %	9	2,5 %	87	23,8 %
QC	3 013	1 900	63,1 %	105	3,5 %	2 005	66,5 %
ONT.	5 634	3 658	64,9 %	155	2,8 %	3 813	67,7 %
MAN.	820	376	45,9 %	53	6,5 %	429	52,3 %
SASK.	808	368	45,5 %	45	5,6 %	413	51,1 %
ALB.	1 880	935	49,7 %	75	4,0 %	1 010	53,7 %
C.-B.	2 020	1 208	59,8 %	101	5,0 %	1 309	64,8 %
YK	29	5	17,2 %	0	0,0 %	5	17,2 %
T. N.-O.	52	15	29,8 %	2	3,8 %	17	32,7 %
NU	41	18	43,9 %	6	14,6 %	24	58,5 %
Canada	15 541	8 865	57,0 %	573	3,7 %	9 438	60,7 %

Afin d'évaluer l'utilité d'une question particulière à des fins analytiques, il était aussi utile de regarder la non-réponse ou le degré de réponse à chaque question pour l'ensemble des répondants. Les tableaux 6 et 7 indiquent les taux de non-réponse pour la plupart des questions après le traitement de l'ensemble de données et la suppression des données incompatibles ou aberrantes.

De plus, quelques questions filtre permettaient de déterminer comment le répondant avait cheminé dans le questionnaire. Par exemple, pour établir si une école aurait dû répondre à la question 43 sur le soutien technique, nous devons d'abord savoir si l'école avait indiqué au moins une heure de soutien technique à la question 42 (qui sert de filtre dans

cet exemple). Vous trouverez aux tableaux 8 et 9 les taux de non-réponse pour ces questions filtre.

Tableau 6  
**Taux de non-réponse aux questions critiques de l'ETICE**

Question	Non-répondants	Taux de non-réponse
2	94	1,4 %
3	0	0,0 %
6	33	0,5 %
15	109	1,6 %
16	315	4,7 %
49	170	2,5 %

*Nota : Questions critiques : 6 676 répondants*

Tableau 7  
**Taux de non-réponse aux questions non critiques de l'ETICE dont les taux de réponse ne dépendent pas de réponses à des questions filtre dans le questionnaire**

Question	Non-répondants	Taux de non-réponse	Question	Non-répondants	Taux de non-réponse
4	361	5,9 %	32	430	7,0 %
5	607	<b>9,9 %</b>	35	419	6,9 %
7	20	0,3 %	37	226	3,7 %
11	3 023	<b>49,5 %</b>	38	362	5,9 %
12	160	2,6 %	39	202	3,3 %
14	124	2,0 %	40	4 008	<b>65,7 %</b>
17	721	<b>11,8 %</b>	41	628	<b>10,3 %</b>
18	169	2,8 %	42	579	<b>9,5 %</b>
19	205	3,4 %	44	242	4,0 %
20	479	<b>7,8 %</b>	45	299	4,9 %
21	210	3,4 %	46	384	6,3 %
22	254	4,2 %	47	250	4,1 %
27	259	4,2 %	48	258	4,2 %

*Nota :*

*Questions sans filtre, non critiques : 6 103 répondants*

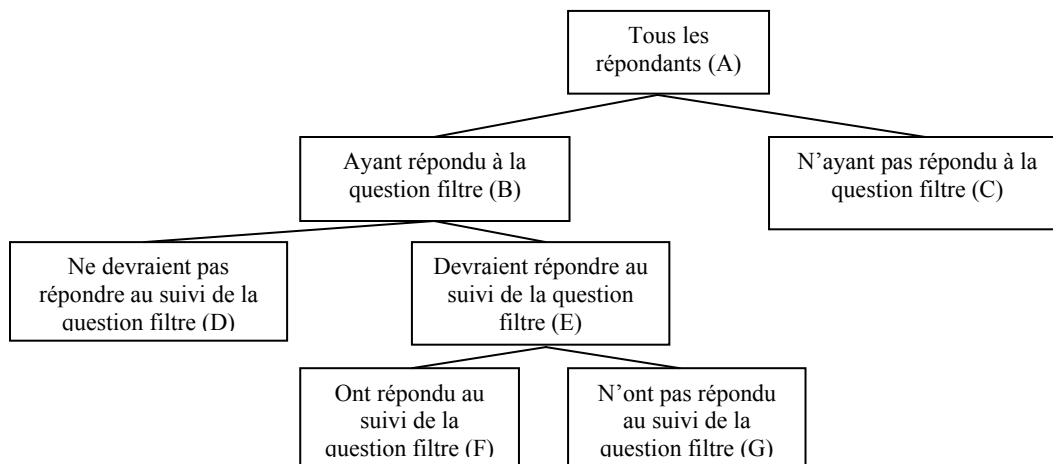
*Les taux de non-réponse élevés sont en gras – seuil arbitraire établi à 7,5 %*

Évidemment, en raison de leur faible taux de réponse, les résultats dérivés des questions 11 et 40 seront beaucoup moins fiables que ceux des autres questions. Les résultats dérivés des autres questions, en gras dans le tableau 7, pourraient aussi ne pas être fiables.

Pour les questions restantes, le taux de non-réponse est indiqué comme un intervalle au lieu d'une valeur fixe, car la non-réponse à la question filtre appropriée brouille le véritable degré de la qualité des données. On cherche à savoir si le répondant qui a répondu à une question filtre aurait dû ou non répondre aux questions suivantes. Toutefois, si une école n'a fourni de donnée d'aucune sorte à la question filtre, alors nous ne savons pas si ces écoles auraient dû répondre aux questions de suivi correspondantes.

Diagramme 1

**Répartition des répondants à une question filtre**



Lorsque nous essayons de déterminer le taux de réponse pour le suivi à une question filtre, nous ne savons pas comment les répondants de l'encadré se seraient répartis entre les encadrés (B), (E) et (G) s'ils avaient répondu à la question filtre.

Au tableau 8, les colonnes « question filtre répondue », « question filtre non répondue » et « non-répondants au suivi » correspondent respectivement aux encadrés (E), (C) et (G) du diagramme 1. De gauche à droite, les trois taux de non-réponse à la droite du tableau sont les limites inférieure et supérieure du taux de non-réponse, de même que la valeur prévue de ce taux. On exprime ces taux par la formule  $[(G) + (X)]/[(E) + (X)]$ , où (E) et (G) désignent le nombre de répondants dans les encadrés correspondants du diagramme 1, tandis que (X) désigne le nombre de répondants dans l'encadré (C) desquels on attendait une réponse à la question de suivi s'ils avaient donné la réponse appropriée à la question filtre.

La limite inférieure du taux de non-réponse s'applique dans le cas où (X) est 0, alors que la limite supérieure s'applique dans le cas où (X) correspond à (C) au complet. La valeur prévue du taux de non-réponse se calcule en prenant d'abord la proportion (E)/(B) des répondants qui ont répondu à la question filtre et qui auraient dû répondre à la question de suivi, en supposant, alors que cette proportion de répondants en (C) aurait aussi été prise en compte. On présume ici que la non-réponse à la question filtre se produit de façon aléatoire en ce qui touche la véritable répartition pour la question filtre.

Tableau 8

**Taux de non-réponse aux questions filtre non critiques de l'ETICE**

Question	Question filtre	Question filtre répondue (E)	Question filtre non répondue (C)	Non-répondants au suivi (G)	Limite inférieure du taux de n.-r.	Limite supérieure du taux de n.-r.	Taux de non-réponse prévu
8	7	5 726	14	297	5,2 %	5,4 %	5,4 %
9	7	5 726	14	734	<b>12,8 %</b>	<b>13,0 %</b>	<b>13,0 %</b>
10	7	5 726	14	320	5,6 %	5,8 %	5,8 %
23	Cellules	189	36	35	<b>18,5 %</b>	<b>31,6 %</b>	<b>19,0 %</b>
24	Internet	5 878	36	333	5,7 %	6,2 %	6,2 %
25	aux	5 878	36	325	5,5 %	6,1 %	6,1 %
26	questions 15 et 16	5 878	36	203	3,5 %	4,0 %	4,0 %
28	27	4 486	203	715	<b>15,9 %</b>	<b>19,6 %</b>	<b>18,7 %</b>
29	27	4 486	203	128	2,9 %	7,1 %	6,1 %
30	27	4 486	203	122	2,7 %	6,9 %	6,0 %
31	27	4 486	203	208	4,6 %	<b>8,8 %</b>	<b>7,8 %</b>
33	32	441	430	103	<b>23,4 %</b>	<b>61,2 %</b>	<b>28,8 %</b>
34	32	441	430	153	<b>34,7 %</b>	<b>66,9 %</b>	<b>39,3 %</b>
36	35	950	349	64	6,7 %	<b>31,8 %</b>	<b>12,1 %</b>
43	42	4 679	579	37	0,8 %	<b>11,7 %</b>	<b>10,2 %</b>

Nota :

Non-réponse aux questions non critiques liées à une question filtre

Les chiffres en gras sont des taux de non-réponse élevés

La question 13 est la seule question dont le taux de non-réponse n'a pas été présenté dans le tableau précédent. Tandis que la première partie de la question s'adresse à tous les répondants, la deuxième partie ne devrait être répondue que par les écoles qui ont des ordinateurs portatifs et des blocs-notes; par conséquent, la cellule c0585 (nombre d'ordinateurs portatifs ou blocs-notes à cette école) peut être considérée comme un filtre pour ces six dernières cellules (c0428-c0433). On présente les taux de non-réponse pour les quatre parties de la question 13 au tableau 9. Les 6 103 répondants de l'étape 2 pouvaient répondre aux deux premières sections, tandis que seulement 2 861 ont clairement indiqué à la question 16 qu'ils possédaient des ordinateurs portatifs ou des blocs-notes, ce qui les rendaient admissibles à répondre à la deuxième partie de la question 13. Par contre, 84 autres répondants n'ont pas indiqué s'ils possédaient ou non des ordinateurs portatifs ou des blocs-notes.

Tableau 9

**Taux de non-réponse à la question 13 par type d'ordinateur**

Type d'ordinateur	Question filtre répondue	Question filtre non répondue	Non-répondants au suivi	Limite inférieure du taux de n.-r.	Limite supérieure du taux de n.-r.	Taux de non-réponse prévu
Odateurs branchés à Internet	6 103	----	137	2,2 %	----	----
Odateurs non branchés à Internet	6 103	----	300	4,9 %	----	----
Ordinateurs portatifs et blocs-notes branchés à Internet	2 984	84	1 089	36,5 %	38,2 %	37,4 %
Ordinateurs portatifs et blocs-notes non branchés à Internet	2 984	84	1 132	37,9 %	39,6 %	38,8 %

*Nota :*

*Non-réponse à la question 13 – la deuxième partie seulement de la question était liée à une question filtre*

## 12.4 Saisie des données

La saisie des données de l'ETICE est décrite au chapitre 8.0. Tel qu'indiqué, on a effectué un contrôle qualitatif qui consistait à comparer la donnée entrée avec la donnée du questionnaire lui-même afin de vérifier l'efficacité du processus de saisie des données. Le processus de contrôle qualitatif a permis de déceler une erreur systématique dans le processus de saisie des données et de la corriger avant l'analyse des données.

## 12.5 Vérification des données et repérage des valeurs aberrantes

On a décrit brièvement au chapitre 8.0 les processus de vérification des données et de repérage des valeurs aberrantes. Dans l'ensemble, 2 951 enregistrements ont déclenché au moins une des 32 vérifications. Les vérifications déclenchées par le plus grand nombre de répondants sont les suivantes (nota : les pourcentages ci-dessous sont les pourcentages pour tous les répondants de l'étape 2) :

- 14,0 % (857) ont indiqué à la question 13 la présence d'ordinateurs portatifs ou blocs-notes dans l'école mais n'en ont pas précisé le nombre à la question 16.
- 6,8 % (418) ont indiqué les pourcentages d'ordinateurs fonctionnant sous différents systèmes d'exploitation, mais ces pourcentages ne totalisaient pas 100 % (question 20).



- 6,7 % (408) n'ont pas indiqué d'heures-personnes par mois affectées au soutien technique à la question 42, mais à la question 43, ils ont donné leurs commentaires au sujet du soutien technique qu'ils recevaient.
- 5,8 % (355) ont indiqué un nombre de comptes de courriel fournis aux élèves (question 25 – c0801) plus élevé que le nombre d'inscriptions à l'école (question 3).

Aucune autre vérification de l'étape 2 n'a été déclenchée par plus de 4,2 % des répondants de l'étape 2. En ce qui concerne les vérifications de l'étape 1, aucune d'entre elles n'a été déclenchée par plus de 3 % (156) de tous les répondants.

Comme nous l'avons dit à la section 8.2, le nombre de suivis effectués auprès des écoles pour corriger les valeurs aberrantes déclarées a été grandement limité par les échéances imposées après la période de collecte. Pour cette raison, nous n'avons pu étendre la vérification des rapports aberrants entre certaines variables numériques. Idéalement, nous aurions contacté un plus grand nombre d'écoles pour confirmer ou corriger les valeurs aberrantes, spécialement celles de la question 2 (où est souvent apparu un nombre élevé d'enseignants) et à la question 4 (où est souvent apparu un nombre élevé d'élèves à temps partiel).

## 13.0 Lignes directrices pour l'analyse, la diffusion et les totalisations

Ce chapitre donne une vue d'ensemble des lignes directrices que doivent observer les utilisateurs lors de la totalisation, de l'analyse, de la publication ou de la diffusion de données provenant du fichier de microdonnées de l'ETICE. Les utilisateurs des microdonnées qui suivront ces lignes directrices devraient pouvoir produire les mêmes statistiques que Statistique Canada.

### 13.1 Lignes directrices pour l'arrondissement

Afin que les estimations destinées à la publication ou à toute autre forme de diffusion qui sont tirées des fichiers de microdonnées de l'Enquête sur les technologies de l'information et des communications dans les écoles (ETICE) correspondent à celles de Statistique Canada, nous conseillons vivement aux utilisateurs de se conformer aux lignes directrices ci-après en ce qui concerne l'arrondissement de ces estimations.

Les estimations dans le corps d'un tableau statistique devraient être arrondies à la centaine près au moyen de la technique d'arrondissement classique. Dans cette technique, si le premier chiffre à supprimer se situe entre 0 et 4, le dernier chiffre à conserver ne change pas. Si le premier ou le seul chiffre à supprimer se situe entre 5 et 9, on augmente de 1 la valeur du dernier chiffre à conserver. Par exemple, lorsqu'on cherche à arrondir à

la centaine près de la façon classique, si les deux derniers chiffres se situent entre 00 et 49, il faut les remplacer par 00 et laisser le chiffre précédent (le chiffre des centaines) tel quel. Si les deux derniers chiffres se situent entre 50 et 99, on les remplace par 00 et on augmente de 1 le chiffre précédent.

Les totaux partiels marginaux et les totaux marginaux des tableaux statistiques devraient être calculés à partir de leurs composantes correspondantes non arrondies, puis arrondis à leur tour à la centaine près à l'aide de la technique d'arrondissement classique.

Les moyennes, les proportions, les taux et les pourcentages devraient être calculés à partir d'éléments non arrondis (c.-à-d. les numérateurs et/ou les dénominateurs), puis arrondis à une décimale au moyen de la technique d'arrondissement classique. Lorsqu'on cherche à arrondir à un seul chiffre par cette technique, si le dernier ou le seul chiffre à supprimer se situe entre 0 et 4, le dernier chiffre à conserver ne change pas. Si le premier ou le seul chiffre à supprimer se situe entre 5 et 9, on augmente de 1 le dernier chiffre à conserver.

Les sommes et les différences d'agrégats (ou de rapports) devraient être calculées à partir de leurs éléments correspondants non arrondis, puis arrondis à leur tour à la centaine près (ou à la décimale près) en employant la technique d'arrondissement classique.

Si, en raison de limitations d'ordre technique ou de toute autre nature, on utilise une autre technique d'arrondissement que l'arrondissement classique, de sorte que les estimations à publier ou à diffuser sous une forme quelconque diffèrent des estimations correspondantes publiées par Statistique Canada, nous conseillons vivement aux utilisateurs d'indiquer la raison de ces divergences dans le ou les documents à publier ou à diffuser.

Des estimations non arrondies ne devraient en aucun cas être publiées ou diffusées par les utilisateurs.

## 13.2 Lignes directrices pour la pondération

Même si l'ETICE est un recensement des écoles primaires et secondaires à travers le Canada et, par conséquent, comporte un plan d'échantillonnage très simple, les non-répondants de l'enquête n'étaient pas répartis également dans la population. Comme nous l'avons décrit au chapitre 10.0, on a choisi un plan de pondération complexe pour bien saisir ce comportement. L'utilisation de données provenant de telles enquêtes complexes pose des problèmes aux analystes, car les procédés nécessaires pour neutraliser la non-réponse touchent et les calculs des estimations ponctuelles et les estimations de variance (voir section 14). Il faut employer des pondérations pour que les estimations ponctuelles soient aussi exemptes que possible de biais.

Pour produire des estimations ponctuelles, y compris des tableaux statistiques ordinaires, les utilisateurs doivent employer la pondération appropriée. Lorsqu'on désire produire des estimations englobant les questions critiques (c.-à-d. les questions 2, 3, 6, 15, 16 et

49), on doit utiliser la pondération **scwght\_p** de la première étape; comme chaque répondant a fourni des données à ces questions, toutes les écoles ayant répondu ont une pondération non nulle pour la première étape. Toutefois, lorsqu'on produit des estimations à partir de questions non critiques, on doit utiliser le poids **scwghtnp** de la deuxième étape; l'ensemble des répondants qui ont refusé de répondre aux questions non critiques affichent une pondération de 0 pour la deuxième étape. Si l'on n'utilise pas les pondérations appropriées, les estimations dérivées des microdonnées ne correspondront pas à celles de Statistique Canada. De plus, les estimations produites à partir des questions non critiques à l'aide des pondérations de l'étape 1 risquent d'être biaisées, car ces pondérations n'ont pas tenu compte de la non-réponse partielle. On peut considérer la pondération attribuée à chaque école comme le nombre d'écoles dans la base de sondage représentées par ce répondant particulier.

Pour certaines techniques d'analyse (par exemple, la régression linéaire, la régression logistique, l'analyse de la variance ou toute autre analyse qui nécessite une mesure significative), on recommande d'utiliser un facteur de pondération corrigé. La méthode utilisée à cette fin rééchelonne les facteurs de pondération de sorte que le poids moyen soit 1 et que la somme de tous les poids corresponde à la taille de l'échantillon.

Supposons, par exemple, qu'il faille analyser toutes les écoles secondaires. Voici les étapes à suivre pour remettre à l'échelle les facteurs de pondération :

- Sélectionner du fichier tous les répondants qui ont été classés comme écoles secondaires (variable GRADLVLD = 2).
- Calculer la pondération moyenne pour ces enregistrements en additionnant les facteurs de pondération finals des écoles (variable scwght\_p ou scwghtnp), à partir du fichier de microdonnées pour ces enregistrements, puis les diviser par le nombre de répondants classés comme écoles secondaires.
- Pour chacun de ces répondants, calculer un facteur de pondération rééchelonné, égal à la pondération finale des écoles et le diviser par le facteur de pondération moyen.
- Effectuer l'analyse, pour ces répondants, à l'aide du facteur de pondération rééchelonné.

Cette méthode peut être utile pour interpréter les résultats provenant d'un logiciel statistique ordinaire. Bien que de nombreuses méthodes d'analyse faisant partie de logiciels statistiques permettent d'utiliser des pondérations, celles-ci diffèrent, sous l'angle de leur définition ou de leur signification, de celles qui conviennent dans le cadre d'une enquête comme l'ETICE. En effet, si les estimations ponctuelles produites au moyen de ces logiciels sont exactes dans bien des cas, les estimations de la variance calculées n'ont à peu près pas de signification. De plus ces estimations pourraient ne pas concorder avec celles qui sont disponibles à Statistique Canada, en raison de la manière

dont certains logiciels statistiques traitent la zone du facteur de pondération. Le rééchantillonnage des facteurs de pondération peut rendre plus significatives les variances dégagées par les logiciels ordinaires en incorporant les probabilités inégales de tirage. Un des avantages de la correction des facteurs de pondération est qu'elle permet d'éviter la surestimation du niveau de signification, lequel dépend fortement de la taille de l'échantillon, tout en conservant les mêmes distributions que celles obtenues au moyen du facteur de pondération final.

Lorsqu'on utilise des logiciels statistiques ordinaires pour calculer des estimations assorties à des facteurs de pondération rééchantillonnés, on ne tient toujours pas compte de la variabilité des estimations inhérentes au mécanisme de non-réponse. C'est pourquoi les estimations de la variance ainsi calculées sous-estiment vraisemblablement la variance véritable. Pour calculer des estimations de la variance plus précises, il faut connaître plus intimement le comportement présumé de la population observée en ce qui concerne la non-réponse de même que le procédé de correction correspondant qui a été choisi. Pour plus de renseignements sur le calcul des estimations de la variance pour l'ETICE, veuillez consulter le chapitre 14.0.

### 13.3 Estimations nominales ou quantitatives

L'unité d'analyse dans le fichier de l'ETICE est l'école. Même si l'on peut tirer d'autres estimations de l'ETICE, les facteurs de pondération de chaque enregistrement (`scwght_p` et `scwghtnp`) ont été conçus pour fournir des estimations fiables de proportions à partir des chiffres des écoles. Par exemple, l'estimation de la proportion d'élèves inscrits dans les écoles qui ont indiqué l'existence d'un site Internet peut ne pas être aussi fiable que l'estimation de la proportion des écoles qui ont déclaré avoir un site Internet.

Avant de discuter de la façon dont on peut totaliser et analyser les données de l'ETICE, il est utile de décrire les deux principaux types d'estimation ponctuelle des caractéristiques de la population que le fichier de microdonnées de l'ETICE permet de produire. Pour simplifier, il faut supposer que l'ensemble des répondants appropriés a été utilisé dans les descriptions suivantes. Cela dépend si l'on a besoin uniquement de variables critiques (répondants de l'étape 1) ou de quelque variable non critique (répondants de l'étape 2). De plus, nous entendrons par **domaine** un groupe d'écoles pour lesquelles on produira une estimation (par exemple, l'ensemble des écoles situées dans une province ou un territoire particulier).

#### *13.3.1 Estimations nominales*

Les estimations nominales sont des estimations du nombre ou du pourcentage de personnes dans la population observée, qui possèdent certaines caractéristiques ou qui font partie d'une catégorie définie. Une école qui utilise la vidéoconférence ou qui a des politiques en place concernant les TIC sont des exemples d'estimations nominales.

Exemples de questions nominales :

Q : Quels types de site Web votre école a-t-elle sur Internet?

R : Site officiel/Site informel/Aucun site Web

Q : Est-ce que votre école a une politique écrite sur l'utilisation acceptable des technologies par les élèves?

R : Oui/Non

### *13.3.2 Estimations quantitatives*

Les estimations quantitatives sont des estimations de totaux ou de moyennes, de médianes ou d'autres mesures de tendance centrale de quantités reposant sur une partie ou la totalité des membres de la population observée. Elles comprennent aussi explicitement des estimations de la forme  $\hat{X} / \hat{Y}$ , où  $\hat{X}$  est une estimation de la quantité totale pour la population observée et  $\hat{Y}$  est une estimation du nombre d'écoles dans la population observée qui contribuent à cette quantité totale, ou d'autres estimations plus complexes comme les coefficients de régression.

Voici un exemple d'estimation quantitative : les dépenses annuelles moyennes par école pour l'enrichissement de la collection de sa bibliothèque. Le numérateur est une estimation des sommes totales consacrées à l'enrichissement de la collection de la bibliothèque, et le dénominateur est une estimation du nombre total d'écoles ayant une bibliothèque.

Exemples de questions quantitatives :

Q : Combien de laboratoires informatiques avez-vous dans votre école?

R :  laboratoires

Q : Combien de dispositifs d'enregistrement numérique (utilisés uniquement à des fins éducatives) avez-vous dans votre école?

R :  dispositifs

### *13.3.3 Totalisation d'estimations nominales*

#### **Nombre d'écoles présentant une caractéristique donnée**

Lorsque tous les répondants du domaine ont répondu à la question liée à la caractéristique qui nous intéresse (c.-à-d. qu'il n'y a aucune non-réponse à une question), on peut obtenir l'estimation à partir du fichier de microdonnées en additionnant les facteurs de pondération finals de tous les enregistrements du domaine affichant cette caractéristique.

S'il y a une non-réponse dans le domaine pour la question pertinente, on peut obtenir l'estimation en deux étapes, telles que décrites ci-après. Cette approche convient lorsque

la non-réponse observée à une question est faible (c.-à-d. moins que le seuil arbitraire de 7,5 % utilisé dans la section 12.3).

Étape 1 : calculer la proportion des écoles du domaine ayant répondu présentant cette propriété (voir le procédé ci-après concernant les proportions)

Étape 2 : multiplier cette proportion par la somme des facteurs de pondération pour toutes les écoles du domaine.

### **Proportions de la forme $\hat{X}/\hat{Y}$**

a) Pour le numérateur ( $\hat{X}$ ), additionner les facteurs de pondération finals des enregistrements présentant la caractéristique qui nous intéresse,

b) obtenir  $\hat{Y}$  en additionnant les facteurs de pondération finals des enregistrements pour toutes les écoles du domaine, en excluant celles qui n'ont pas répondu à la question qui servait à identifier la caractéristique,

c) diviser l'estimation a) par l'estimation b) ( $\hat{X} / \hat{Y}$ ).

### *13.3.4 Totalisation d'estimations quantitatives*

#### **Estimation des chiffres de population**

On peut obtenir des estimations de quantités à partir du fichier de microdonnées en multipliant la valeur de la variable qui nous intéresse par le facteur de pondération final pour chaque enregistrement, lorsque chaque répondant du domaine a fourni le renseignement, puis en additionnant cette quantité pour tous les enregistrements visés. Comme c'était le cas pour les estimations nominales, s'il n'y a pas de non-réponse dans un domaine particulier pour la question pertinente, on doit alors dériver l'estimation en deux étapes, telles que décrites ci-après.

Étape 1 : obtenir la valeur moyenne de la quantité qui nous intéresse pour les écoles ayant répondu dans le domaine en utilisant la procédure ci-après pour les ratios.

Étape 2 : multiplier cette moyenne par le total des facteurs de pondération pour toutes les écoles du domaine.

Si la variable quantitative est corrélée à l'inscription dans l'école, on peut affiner l'estimation en utilisant cette information complémentaire. L'approche suivante ne s'applique que pour les variables corrélées au nombre d'élèves, car il n'y a pas d'autres variables dans cette étude pour lesquelles aucune non-réponse à une question n'a été observée.

Étape 1 : obtenir la valeur moyenne pour la quantité qui nous intéresse par élève dans le domaine en utilisant la procédure ci-après pour calculer les ratios, mais en transformant l'étape b) pour dériver  $\hat{Y}$  en utilisant le nombre d'élèves comme variable d'intérêt.

Étape 2 : faire le total des produits du facteur de pondération et du nombre d'élèves de chaque enregistrement, pour toutes les écoles du domaine.

Étape 3 : multiplier la moyenne obtenue à l'étape 1 par le total de l'étape 2.

Veillez noter que les approches dans cette section ne s'appliquent que si la non-réponse à la question pertinente dans le domaine qui nous intéresse est faible (c.-à-d. moins que le seuil arbitraire de 7,5 % indiqué dans la section 12.3).

### **Proportions et ratios de la forme $\hat{X}/\hat{Y}$**

a) Additionner les produits de la valeur de la variable d'intérêt pour le numérateur et le facteur de pondération final des enregistrements du domaine pour le numérateur ( $\hat{X}$ ), en excluant ceux qui n'ont pas répondu aux variables d'intérêt qui correspondent au numérateur ou au dénominateur.

b) obtenir  $\hat{Y}$  en additionnant ensemble les produits de la valeur de la variable d'intérêt pour le dénominateur et les facteurs de pondération finals des enregistrements pour toutes les écoles du domaine, en excluant ceux qui n'ont pas répondu aux variables d'intérêt qui correspondent au numérateur ou au dénominateur.

c) diviser l'estimation a) par l'estimation b) ( $\hat{X} / \hat{Y}$ ).

Par exemple, pour estimer la proportion des ordinateurs dans les écoles urbaines dotés d'un faible processeur,

- estimer le nombre total d'ordinateurs dans les écoles urbaines dotés d'un faible processeur ( $\hat{X}$ ) en totalisant le produit du nombre d'ordinateurs dotés d'un faible processeur et les facteurs de pondération finals pour les écoles urbaines (dans ce cas-ci on s'attend à ce que toutes les écoles aient répondu aux questions sur le nombre d'ordinateurs),
- estimer le nombre total d'ordinateurs dans les écoles urbaines ( $\hat{Y}$ ) en totalisant le produit du nombre d'ordinateurs et les facteurs de pondération finals de toutes les écoles urbaines,
- diviser la première estimation par la deuxième estimation ( $\hat{X} / \hat{Y}$ ).

### 13.4 Lignes directrices pour la diffusion de coefficients de variation

Avant de diffuser et/ou de publier des estimations tirées de l'ETICE, les utilisateurs devraient commencer par établir le niveau de qualité de l'estimation. Les niveaux de qualité sont acceptable, médiocre et inacceptable. Les erreurs d'échantillonnage et non dues à l'échantillonnage influencent généralement la qualité des données.

Toutefois, comme nous l'avons vu au chapitre 12.0, la qualité d'une estimation dérivée de l'ETICE est basée uniquement sur les erreurs non dues à l'échantillonnage. Un indicateur important de cette qualité est la variabilité liée à la non-réponse et les corrections correspondantes reflétées dans le coefficient de variation, comme le décrit le chapitre 14.0. Vous trouverez au tableau 10 le rapport entre les niveaux types de qualité et le coefficient de variation. Les utilisateurs devraient lire le chapitre 12.0 pour mieux connaître les caractéristiques relatives à la qualité de ces données.

Pour établir le niveau de qualité type d'une estimation, les utilisateurs devraient commencer par établir le nombre de répondants inclus dans le calcul de l'estimation. Si ce nombre est cinq ou moins, l'estimation pondérée ne devrait pas être diffusée afin de se conformer aux politiques de confidentialité. Pour les estimations pondérées reposant sur des tailles d'échantillon supérieures à cinq, les utilisateurs devraient déterminer le coefficient de variation de l'estimation et suivre les lignes directrices ci-après. On doit appliquer ces lignes directrices aux estimations arrondies pondérées. Toute estimation dont le niveau de qualité est médiocre ou inacceptable devrait être accompagnée d'une mise en garde à l'intention des utilisateurs subséquents.

Tableau 10

**Lignes directrices sur le niveau de qualité en fonction du CV d'une estimation particulière**

Niveau de qualité de l'estimation	Lignes directrices
1) Acceptable	Caractéristiques des estimations : Taille de l'échantillon supérieure à cinq et coefficients de variation situés dans un intervalle de 0,0 % à 16,5 %.  Aucune mise en garde n'est requise.
2) Médiocre	Caractéristique des estimations : Taille de l'échantillon supérieure à cinq et coefficients de variation situés dans un intervalle de 16,6 % à 33,3 %.  On doit désigner ces estimations par la lettre M (ou un autre identificateur semblable). Elles doivent faire l'objet d'une mise en garde à l'intention des utilisateurs subséquents au sujet du



---

niveau d'erreur élevé associé aux estimations.

---

3) Inacceptable On ne devrait pas diffuser les estimations associées à une taille de l'échantillon de cinq ou moins, afin de se conformer aux politiques de confidentialité de l'organisme.

Pour des tailles d'échantillon supérieures à cinq mais liées à des coefficients de variation très élevés dépassant 33,3 % %, Statistique Canada recommande de ne pas diffuser d'estimations, en raison de leur niveau de qualité inacceptable. Cependant, si l'utilisateur choisit de le faire, il doit alors désigner les estimations par la lettre I (ou un autre identificateur semblable) et les diffuser avec la mise en garde suivante :

« Veuillez noter que ces estimations [marquées de la lettre I] ne répondent pas aux normes de qualité de Statistique Canada. Les conclusions tirées de ces données ne sauraient être fiables et seront fort probablement erronées. Ces données et toute conclusion qu'on pourrait en tirer ne devraient pas être publiées. Si l'utilisateur choisit de les publier, il est alors tenu de les accompagner de cette mise en garde. »

---

Pour illustrer le concept de coefficients de variation, vous trouverez ci-après un tableau de CV pour une diversité d'estimations de l'ETICE. Veuillez noter que le CV estimé de 0,00 % pour la proportion des écoles du Yukon avec connexion permanente à Internet s'explique par le fait que toutes les écoles participantes avaient indiqué qu'elles étaient branchées à Internet de cette manière. Cela ne signifie pas que toutes les écoles du Yukon qui n'ont pas répondu ont aussi une connexion permanente.

Tableau 11

**Estimations du coefficient de variation pour plusieurs caractéristiques clés par province et territoire, 2003-2004**

	Canada	Terre-Neuve-et-Labrador	Île-du-Prince-Édouard	Nouvelle-Écosse	Nouveau-Brunswick	Québec	Ontario
<b>Proportion d'ordinateurs selon la connectivité à Internet</b> <i>Connectés à Internet</i>	0,18 %	0,62 %	1,38 %	0,36 %	0,24 %	0,52 %	0,32 %

<i>Non connectés à Internet</i>	2,35 %	8,31 %	15,44 %	8,26 %	6,13 %	5,51 %	4,90 %
<b>Proportion d'ordinateurs selon la vitesse du processeur<sup>1</sup></b>							
<i>Faible vitesse</i>	1,25 %	3,59 %	7,03 %	5,05 %	1,91 %	2,26 %	2,71 %
<i>Vitesse moyenne</i>	0,74 %	2,97 %	5,08 %	2,64 %	3,10 %	1,73 %	1,32 %
<i>Haute vitesse</i>	2,12 %	7,60 %	32,88 %	8,27 %	3,12 %	5,80 %	3,99 %
<i>Vitesse inconnue</i>	9,53 %	x	x	29,56 %	18,28 %	31,08 %	19,21 %
<b>Proportion des écoles selon la méthode d'accès à Internet</b>							
<i>Accès commuté<sup>2</sup></i>	3,43 %	9,63 %	x	X	16,18 %	7,02 %	7,44 %
<i>Accès toujours ouvert<sup>3</sup></i>	0,46 %	1,61 %	5,77 %	0,95 %	0,66 %	1,01 %	0,95 %
<b>Ratio du nombre d'élèves par ordinateur (médiane)</b>	0,66 %	3,18 %	4,77 %	1,87 %	1,44 %	1,85 %	0,96 %

	Manitoba	Saskatchewan	Alberta	Colombie-Britannique	Yukon	Territoires du Nord-Ouest	Nunavut
<b>Proportion d'ordinateurs selon la connectivité à Internet</b>							
<i>Connectés à Internet</i>	0,80 %	0,41 %	0,38 %	0,54 %	1,74 %	0,70 %	2,18 %
<i>Non connectés à Internet</i>	6,26 %	6,79 %	4,97 %	5,20 %	x	17,54 %	15,85 %
<b>Proportion d'ordinateurs selon la vitesse du processeur<sup>1</sup></b>							
<i>Faible vitesse</i>	4,20 %	4,93 %	3,07 %	3,02 %	4,83 %	8,44 %	13,60 %
<i>Vitesse moyenne</i>	3,01 %	3,73 %	1,61 %	1,94 %	5,69 %	6,99 %	11,44 %
<i>Haute vitesse</i>	6,25 %	6,30 %	3,43 %	6,07 %	17,04 %	9,19 %	20,82 %
<i>Vitesse inconnue</i>	20,55 %	16,39 %	20,38 %	21,57 %	x	x	x
<b>Proportion des écoles selon la méthode d'accès à Internet</b>							
<i>Accès commuté<sup>2</sup></i>	8,03 %	16,35 %	8,21 %	11,44 %	x	x	x
<i>Accès toujours ouvert<sup>3</sup></i>	1,52 %	1,26 %	1,18 %	1,19 %	0,00 %	6,43 %	22,86 %
<b>Ratio du nombre d'élèves par ordinateur (médiane)</b>	2,60 %	1,72 %	1,32 %	1,45 %	4,28 %	8,12 %	20,62 %

Tableau 12

**Estimations du coefficient de variation pour plusieurs caractéristiques clés par type d'école, 2003-2004**

	Ensemble des écoles	Niveau d'enseignement de l'école		Emplacement de l'école	
		Élémentaire	Secondaire	Urbain	Rural
<b>Proportion d'ordinateurs selon la connectivité à Internet</b>					
<i>Connectés à Internet</i>	0,18 %	0,25 %	0,30 %	0,21 %	0,34 %
<i>Non connectés à Internet</i>	2,35 %	2,55 %	5,97 %	2,80 %	3,86 %
<b>Proportion d'ordinateurs selon la vitesse du processeur<sup>1</sup></b>					
<i>Faible vitesse</i>	1,25 %	1,36 %	2,60 %	1,47 %	2,18 %
<i>Vitesse moyenne</i>	0,74 %	0,95 %	1,26 %	0,84 %	1,47 %
<i>Haute vitesse</i>	2,12 %	2,56 %	3,52 %	2,45 %	3,93 %
<i>Vitesse inconnue</i>	9,53 %	10,00 %	22,90 %	10,78 %	18,36 %
<b>Proportion des écoles selon la méthode d'accès à Internet</b>					
<i>Accès commuté<sup>2</sup></i>	3,43 %	4,43 %	11,59 %	6,12 %	4,08 %

Accès toujours ouvert <sup>3</sup>	0,46 %	0,61 %	0,67 %	0,48 %	1,20 %
Ratio du nombre d'élèves par ordinateur (médiane)					
Canada	0,66 %	0,73 %	1,17 %	0,68 %	1,34 %

## 14.0 Calcul de la variance

Même si l'ETICE n'est pas une enquête probabiliste, en ce sens qu'aucun échantillon n'a été tiré pour représenter la population, il y a toujours une certaine variabilité dans les estimations en raison des erreurs non dues à l'échantillonnage. Une part importante de cette variabilité tient aux schémas de réponse et aux corrections visant à tenir compte des éléments de non-réponse. Ce chapitre explique pourquoi il est important de calculer la variance et présente différents outils pour le faire.

### 14.1 Importance de la variance

La variance d'une estimation est une bonne indication de la qualité d'une estimation. Une estimation assortie d'une variance trop élevée est considérée comme non fiable. Pour connaître le degré de variance, on utilise une mesure relative de la variabilité, le coefficient de variation, ou CV.

Le coefficient de variation se définit comme le ratio de la racine carrée de la variance (aussi désignée comme l'écart-type ou l'erreur-type) à l'estimation. L'utilisation du coefficient de variation plutôt que la variance permet à l'analyste de comparer sur une même échelle des estimations de grandeurs différentes. On peut donc quantifier la qualité de toute estimation au moyen du coefficient de variation.

Il est possible de déterminer la véritable variance d'une estimation, et par le fait même le véritable CV, seulement si nous connaissons la caractéristique correspondante qui nous intéresse pour toutes les écoles de la population visée. L'autre possibilité, si tel est le cas, est d'estimer la variance en fonction de l'ensemble des répondants. Pour simplifier la discussion dans ce document, la variance et le CV dans ce chapitre seront désignés par leurs valeurs estimées.

#### **Le rôle des questions critiques dans l'estimation de la variance**

La méthode utilisée pour l'approximation de la variance de l'estimation a été différente selon que l'estimation s'appliquait aux questions critiques (c.-à-d. questions 2, 3, 6, 15, 16 et 49) ou aux questions non critiques. Lorsque seules les questions critiques étaient nécessaires pour produire une estimation, l'estimation de l'erreur-type était calculée en une seule étape, d'après l'ensemble des répondants à l'enquête. Cependant, lorsque les questions non critiques étaient utilisées, il fallait procéder en deux étapes. La première étape englobait tous les répondants et la seconde, seulement un sous-ensemble de répondants qui avaient répondu au-delà des questions critiques. Par exemple, l'erreur-type liée au nombre moyen d'ordinateurs connectés à Internet par école (questions 15 et

16) a été obtenue en une seule étape. Toutefois, l'erreur-type associée à la proportion d'écoles avec un ou des logiciels de traitement de texte mis à la disposition des élèves (question 18) a nécessité une approche en deux étapes.

### **Méthode pour obtenir la variance d'une estimation**

À partir des hypothèses émises au sujet du mécanisme de non-réponse décrit au chapitre 10.0, il est possible de dériver une formule exacte pour estimer la variance pour chaque estimation ponctuelle dans l'ETICE. Toutefois, il serait extrêmement coûteux de procéder ainsi pour chacune des estimations susceptible de provenir de l'ETICE et une telle procédure serait inutile pour la plupart des utilisateurs.

Il existe un outil convivial pour aider les utilisateurs à calculer la variance et le CV de leurs estimations. Cet outil est un module de type Excel pour l'extraction du CV (MECV) qui produit des CV approximatifs pour un grand nombre de domaines. Le MECV a été conçu principalement pour les proportions même s'il peut fournir des approximations grossières de CV pour d'autres statistiques simples (p. ex., totaux pour des variables nominales).

Il est aussi important de souligner que tous les coefficients de variance produits par le MECV sont approximatifs et ne doivent pas être considérés comme des valeurs officielles. Moyennant des frais, on peut se procurer des estimations des CV réels pour des variables particulières en s'adressant à Statistique Canada. L'utilisation d'estimations officielles de CV peut permettre aux utilisateurs de diffuser des estimations que l'on ne pourrait diffuser autrement, c'est-à-dire des estimations avec des CV de type « restreint ».

## **14.2 Module de type Excel pour l'extraction de coefficients de variation**

Le module de type Excel pour l'extraction de CV (MECV), conçu à l'aide de macros Excel et utilisable grâce à une interface conviviale, a d'abord été créé pour l'extraction de CV d'estimations de proportion provenant de variables nominales de l'ETICE. Même si l'outil peut aussi servir à extraire des CV pour des estimations de totaux liées à des variables nominales, leur fiabilité sera moindre que pour les CV de proportion, sauf dans le cas spécial où la variable correspondante ne présente aucune non-réponse à une question. On peut extraire des CV de deux façons, décrites ci-après.

La première façon consiste à définir le domaine d'intérêt à l'aide des cinq variables analytiques suivantes : la province/territoire (ProvCode) où se trouve l'école, le type d'école, urbaine ou rurale (RuralDChosen), le type de financement de l'école (TypeSchD), le niveau d'enseignement (GradLvID) et la taille (SchSizeD) de l'école. La deuxième façon demande de préciser la taille du domaine d'intérêt. L'information comprend : l'estimation de la proportion, le nombre de répondants dans le domaine spécifié, la population estimée dans ce domaine, les statistiques élémentaires (la variance, l'erreur-type et l'intervalle de confiance à 95 % correspondant à l'estimation ponctuelle de la population) et le coefficient de variation pour la proportion sélectionnée.

Le MECV couvre plus de 200 domaines. À l'échelon national, les utilisateurs peuvent extraire des CV de n'importe quel domaine englobant au plus deux variables analytiques (par exemple, par niveau d'enseignement, ou par taille de l'école et financement, mais non par financement, niveau d'enseignement et emplacement). À l'échelon provincial/territorial, on peut extraire des CV de n'importe quel domaine incluant au plus une de ces variables (p. ex., pour la province du Manitoba ou pour la province de Québec et le financement, mais non pour la province de l'Alberta et le niveau d'enseignement et la taille de l'école). On a produit un CV approximatif pour huit proportions différentes dans chacun des domaines, pour un total de plus de 1 600 CV. Toutefois, on a supprimé les CV des domaines dans lesquels cinq enregistrements ou moins avaient contribué à la proportion en question.

Nous avons utilisé des simulations, basées sur 100 répétitions et une hypothèse de non-réponse à une question, pour estimer les variances, les coefficients de variation et les intervalles de confiance à 95 % pour différentes proportions, c.-à-d. 1 %, 5 %, 10 %, 15 %, 20 %, 30 %, 40 % et 50 %. Afin d'estimer le CV pour une proportion supérieure à 50 %, l'utilisateur peut obtenir le CV pour la proportion de 50 % et présumer que le CV désiré sera moindre que cette quantité. Il peut aussi obtenir un CV pour une proportion dépassant 50 % en tenant pour acquis que l'erreur-type pour une proportion  $p$  est égale à l'erreur-type pour la proportion correspondante  $(1-p)$ . Par exemple, si le MECV indique un CV de 0,06 pour une proportion  $p$ , le CV pour  $(1-p)$  sera alors

$$CV = 0,06 * (p)/(1-p)$$

Comme nous l'avons mentionné au chapitre 13.0, les utilisateurs devraient toujours vérifier la qualité des estimations, particulièrement pour les proportions plus petites obtenues de petits domaines. Un code de couleur dans l'application Excel indique les CV élevés. Les CV dont l'intervalle dépasse 50 % apparaissent en rouge, ceux dont l'intervalle est situé entre 33,3 % et 50 % s'affichent en orange et ceux dont l'intervalle se situe entre 16,6 % et 33,3 % sont en jaune. Pour plus de renseignements au sujet des lignes directrices pour la diffusion des coefficients de variation pour l'ETICE, veuillez consulter la section 13.5. D'autres renseignements à l'égard des lignes directrices sur la qualité pour les produits de Statistique Canada figurent dans la publication intitulée : *Les lignes directrices concernant la qualité de Statistique Canada*, n° 12-539-XIF au catalogue.

Des règles s'appliquent si l'on désire dériver des CV pour des estimations nominales à l'aide du MECV.

### **Règles 1 : Estimations du nombre d'écoles présentant une caractéristique donnée (agrégats)**

Étant donné que le MECV ne tient pas compte de la non-réponse à une question, on peut affirmer sans crainte que le CV d'une estimation se rapproche de celui de la proportion qu'il représente (tout en lui étant légèrement supérieur) dans la mesure où la non-réponse

à une question correspondante est faible. Donc, pour obtenir le CV approximatif d'une estimation, on peut utiliser le MECV en y précisant la taille du domaine et en dérivant la proportion appropriée. On devrait employer cette procédure uniquement quand la non-réponse à la question liée à la caractéristique est faible (moins que le seuil arbitraire de 7,5 % à la section 12.3) et que la taille du domaine d'intérêt est relativement élevée. Généralement, les estimations provinciales/territoriales seront fiables, mais les estimations infraprovinciales/infraterritoriales n'auront de fiabilité que dans la mesure où le domaine est apparenté aux groupes de correction de la non-réponse ou à l'ensemble des totaux utilisés pour calibrer les pondérations.

Par exemple, supposons que nous avons une estimation  $\hat{Y} = 30\ 000$  écoles présentant une caractéristique donnée. Si nous voulons les comparer aux  $100\ 000$  écoles du domaine d'intérêt, alors le CV pour  $\hat{Y}$  devrait se rapprocher du CV de la proportion (qui est, dans notre exemple,  $30\ 000 / 100\ 000 = 30,0\ %$ ).

### **Règle 2 : Estimations de proportions ou de pourcentages d'écoles présentant une caractéristique donnée**

Les CV calculés dans le MECV le sont pour des proportions. Ils peuvent donc être utilisés tels qu'ils apparaissent dans le chiffrier. Toutefois, ils ne sont valides que si la non-réponse à une question est faible (moins que le seuil arbitraire établi de 7,5 % utilisé dans la section 12.3). Par conséquent, il n'est pas recommandé d'utiliser le MECV pour estimer le CV de proportions reposant sur des données des questions 11 ou 40.

### **Règle 3 : Estimations des différences entre des agrégats, des pourcentages et des ratios**

L'erreur-type d'une différence entre deux estimations est à peu près égale à la racine carrée de la somme des carrés de chaque erreur-type considérée séparément.

L'approximation de l'erreur-type de différence ( $\hat{d} = \hat{X}_1 - \hat{X}_2$ ) est :

$$\sigma_{\hat{d}} = \sqrt{(\hat{X}_1\alpha_1)^2 + (\hat{X}_2\alpha_2)^2}$$

où  $\alpha_1$  et  $\alpha_2$  sont respectivement les CV de  $\hat{X}_1$  et  $\hat{X}_2$ . On obtient le CV de  $\hat{d}$  en divisant son erreur-type par  $\hat{d}$  elle-même. C'est une bonne approximation du CV d'une différence entre des caractéristiques distinctes et non corrélées, mais elle peut donner des résultats inexacts dans les autres cas, car elle tendra à surestimer le CV si  $\hat{X}_1$  et  $\hat{X}_2$  sont corrélés positivement et à sous-estimer le CV si  $\hat{X}_1$  et  $\hat{X}_2$  sont corrélés négativement. Moyennant des frais, on peut obtenir les estimations de CV pour les valeurs  $\hat{X}_1$  et  $\hat{X}_2$  corrélées en s'adressant à Statistique Canada.

#### **Règle 4 : Estimations de ratios**

Dans le cas spécial où le ratio représente une proportion ou un pourcentage des écoles présentant une caractéristique donnée, on peut alors utiliser le MECV comme dans la règle 2. Autrement, l'erreur-type du ratio des estimations est approximativement égale à la racine carrée de la somme des carrés de chaque CV pris séparément, multipliées par le ratio lui-même. Le CV d'un tel ratio  $\hat{R} = \hat{X}_1 / \hat{X}_2$  est

$$\alpha_{\hat{R}} = \sqrt{\alpha_1^2 + \alpha_2^2}$$

où  $\alpha_1$  et  $\alpha_2$  sont les CV de  $\hat{X}_1$  et  $\hat{X}_2$  respectivement. Comme dans l'estimation du CV de la règle 3, cette estimation tend à surestimer le CV si  $\hat{X}_1$  et  $\hat{X}_2$  sont corrélés positivement et à sous-estimer le CV si  $\hat{X}_1$  et  $\hat{X}_2$  sont corrélés négativement.

#### **Règle 5 : Estimations des différences entre des ratios**

Il s'agit ici simplement de combiner les règles 3 et 4. On commence par calculer le CV des deux ratios au moyen de la règle 4, puis on trouve le CV de leur différence par la règle 3.

### 14.3 Calculer les limites de confiance à l'aide du coefficient de variation

Bien que les coefficients de variation soient largement utilisés, l'intervalle de confiance d'une estimation donne une mesure intuitive plus significative de l'erreur d'échantillonnage. Un intervalle de confiance est le niveau de confiance que la valeur vraie de la population se situe dans une fourchette donnée de valeurs. Dans le cas de l'ETICE, où l'on a utilisé un recensement plutôt qu'un échantillonnage pour contacter les répondants potentiels, la variabilité dans les estimations dépend premièrement de la non-réponse observée. Comme les corrections pour la non-réponse ont été apportées en fonction du nombre de répondants dans chaque catégorie de correction, l'intervalle de confiance à 95 % signifie que si le recensement se répète à l'infini et que dans chaque cas le nombre de répondants dans chaque catégorie de correction demeure constant, 95 % des ensembles de réponses produiront alors des estimations dont l'intervalle de confiance contiendrait la vraie valeur.

Les intervalles de confiance pour une estimation,  $\hat{X}$ , sont généralement exprimés sous forme de deux nombres, l'un inférieur et l'autre supérieur à l'estimation, comme dans  $(\hat{X} \pm k)$ , où  $k$  varie selon le niveau de confiance souhaité et l'erreur d'échantillonnage de l'estimation. Comme la véritable erreur-type des estimations est inconnue, les intervalles de confiance sont estimés en fonction de l'ensemble des répondants.

Une fois déterminée l'erreur-type d'une estimation, on peut calculer des intervalles de confiance pour les estimations en partant de l'hypothèse qu'avec un échantillon

suffisamment vaste, les diverses estimations obtenues pour une caractéristique de la population sont réparties selon une distribution normale autour de la vraie valeur de la population. Selon cette hypothèse, il y a environ 68 chances sur 100 que la différence entre une estimation de l'échantillon et la vraie valeur pour la population soit inférieure à une erreur-type, environ 95 chances sur 100 que la différence soit inférieure à deux erreurs-types et environ 99 chances sur 100 que la différence soit inférieure à trois erreurs-types. On appelle ces différents degrés de confiance des niveaux de confiance.

Les intervalles de confiance à 95 % pour une estimation sont disponibles directement dans le chiffrier des CV. Si l'utilisateur veut déterminer d'autres intervalles de confiance, il peut employer la formule ci-après pour obtenir l'intervalle de confiance (CI  $\hat{X}$ ) :

$$CI \hat{X} = ( \hat{X} \pm z\hat{X}\alpha_{\hat{X}} )$$

où  $\alpha_{\hat{X}}$  est le CV trouvé pour  $\hat{X}$  et

$z = 1$  si l'on désire un intervalle de confiance à 68 %;

$z = 1,65$  si l'on désire un intervalle de confiance à 90 %;

$z = 1,96$  si l'on désire un intervalle de confiance à 95 %;

$z = 2,58$  si l'on désire un intervalle de confiance à 99 %.

### **Exemple d'utilisation du coefficient de variation pour obtenir des limites de confiance**

Un intervalle de confiance à 95 % pour une proportion estimée d'écoles élémentaires utilisant un modem pour accéder à Internet s'obtiendrait de la façon suivante :

$$\hat{X} = 16,7 \% \text{ (ou, exprimé comme une proportion, } 0,166635)$$

$$z = 1,96$$

Le coefficient de variation de  $\hat{X}$  est 3,14 % (ou 0,0314, lorsqu'il est exprimé comme une proportion). C'est le coefficient de variation de l'estimation obtenu à partir du module d'extraction des CV.

$$IC \hat{X} = 0,167 - (1,96) (0,167) (0,0314), 0,167 + (1,96) (0,167) (0,0314)\}$$

$$IC \hat{X} = 0,167 - 0,010, 0,167 + 0,010\}$$

$$IC \hat{X} = 0,157, 0,177\}$$

On peut donc dire, avec un niveau de confiance à 95 %, qu'entre 15,9 % et 17,5 % des écoles élémentaires utilisent un modem pour accéder à Internet.

## 14.4 Tests d'hypothèses

On peut aussi utiliser les erreurs-types pour vérifier des hypothèses, une technique qui permet de distinguer les paramètres d'une population à l'aide d'estimations basées sur un



échantillon. Ces estimations peuvent être des nombres, des moyennes, des pourcentages, des ratios, etc. Les tests peuvent être effectués à divers niveaux de signification, où un niveau de signification est la probabilité de conclure que les caractéristiques sont différentes quand, en fait, elles sont identiques.

Posons  $\hat{X}_1$  et  $\hat{X}_2$  comme des estimations d'échantillon pour deux caractéristiques d'intérêt. Comme dans la règle 3 à la section 14.2, supposons que l'erreur-type de la différence  $\hat{X}_1 - \hat{X}_2$  est  $\sigma_d$ . Si la quantité  $(\hat{X}_1 - \hat{X}_2) / \sigma_d$  se situe entre  $-1,96$  et  $+1,96$ , on ne peut alors tirer de conclusions à propos de la différence entre les caractéristiques au niveau de signification de 5 %. Toutefois, si ce rapport est inférieur à  $-1,96$  ou supérieur à  $+1,96$ , la différence observée est significative au niveau de 5 %. Comme l'indique la règle 3 à la section 14.2, cette méthode est fiable si elle s'applique à deux caractéristiques d'intérêt non corrélées – autrement, la quantité  $(\hat{X}_1 - \hat{X}_2) / \sigma_d$  peut s'avérer inexacte.

On peut aussi déterminer si deux estimations sont significativement différentes en comparant les intervalles de confiance des deux estimations. Lorsque deux intervalles de confiance se chevauchent, on peut conclure qu'il n'y a pas de différence significative entre les estimations correspondantes. Il convient de souligner que cette méthode est plutôt approximative, car elle revient à approximer l'intervalle de confiance pour la différence des deux estimations. Par conséquent, l'approximation de l'intervalle de confiance sera légèrement supérieure à celle qui est dérivée dans la section 13.3. Pour cette raison, le recours à cette règle empirique prudente entraîne un faible risque de ne pas détecter certains écarts significatifs (mais très minces) selon le niveau de confiance utilisé.

#### 14.5 Coefficients de variation d'estimations quantitatives

Pour calculer le coefficient de variation d'estimations quantitatives, il faudrait élaborer des tables spéciales, ce qui n'a pas été fait puisqu'un grand nombre des variables de l'ETICE sont surtout de type nominal.

Néanmoins, le coefficient de variation d'un total quantitatif sera généralement supérieur à celui de l'estimation nominale correspondante (c.-à-d. l'estimation du nombre d'écoles contribuant à l'estimation quantitative). Si l'estimation nominale correspondante ne peut être diffusée, il en sera de même pour l'estimation quantitative. Par exemple, le coefficient de variation du nombre total d'employés des écoles affectés à la bibliothèque des écoles serait plus grand que le coefficient de variation de la proportion correspondante des écoles ayant une bibliothèque. Par conséquent, si le coefficient de variation de la proportion ne peut être diffusé, ce sera aussi le cas pour le coefficient de variation de l'estimation quantitative correspondante.

Moyennant des frais, les utilisateurs désireux de calculer les coefficients de variation d'estimations quantitatives peuvent demander à Statistique Canada de produire des estimations de CV exactes.

## Annexe A : Questionnaire

Cette annexe présente le questionnaire de l'ETICE envoyé aux directeurs d'école.

Ce questionnaire est en français. Toutefois, sa version anglaise peut être fournie sur demande.

## Annexe B : Caractéristiques d'une école

Cette annexe présente une vue d'ensemble de définitions utilisées pour dériver les variables des caractéristiques d'une école.

### **Niveau d'enseignement de l'école** – Écoles primaires, secondaires et primaires-secondaires mixtes

Les écoles au Canada sont classées comme écoles **primaires** si elles offrent un enseignement de la première à la sixième année ou une majorité des années du niveau primaire; comme écoles **secondaires** si elles offrent un enseignement à partir de la septième année ou une majorité des années du niveau secondaire, et comme **écoles mixtes primaires-secondaires** si elles offrent une combinaison d'années des niveaux primaire et secondaire.

### **Emplacement de l'école** : Écoles urbaines et rurales

Les écoles **rurales** s'entendent des établissements situés en milieu rural, dans les petites villes de même que dans les banlieues rurales d'une agglomération plus importante (régions métropolitaines de recensement (RMR) ou agglomérations de recensement (AR)). Les écoles **urbaines** s'entendent des établissements situés dans une RMR ou une AR, mais non dans les banlieues rurales.

### **Financement de l'école** – Financement public et privé

On a classé les établissements en écoles publiques ou privées selon que la prise de décision quant aux affaires de l'établissement relève, en dernière instance, d'un organisme public ou privé. Sont classés dans la catégorie des écoles **publiques** les établissements qui, selon les déclarations du directeur, sont : contrôlés ou administrés directement par une autorité scolaire ou un organisme public; ou contrôlés et administrés soit directement par un organisme public, soit par un organe directeur (conseil, comité, etc.), dont la majorité des membres est nommée par des pouvoirs publics ou élus par le public. Sont classés dans la catégorie des écoles **privées** les établissements qui, selon les déclarations du directeur, sont contrôlés et administrés par un organisme non gouvernemental (p. ex., une église, un syndicat ou une entreprise) ou dont le conseil d'administration est majoritairement formé de membres non nommés par un organisme public.

### **Taille de l'école** – Petite, moyenne et grande

La taille de l'école est déterminée en fonction de la répartition du nombre d'élèves inscrits dans des écoles primaires, secondaires et mixtes. Les écoles dont le nombre d'élèves se situe dans le tiers inférieur de la distribution (<33,3 %) sont considérées

comme de « **petites écoles** ». Les « **écoles moyennes** » sont celles dont le nombre d'élèves se situe dans le tiers intermédiaire de la distribution ( $\leq 33,3\%$  et  $\leq 66,6\%$ ). Les « **grandes écoles** » s'entendent des établissements dont le nombre d'élèves se situe dans le tiers supérieur de la distribution ( $> 66,6\%$ ). La taille des écoles est donc définie comme suit :

	<b>Primaire</b>	<b>Secondaire</b>	<b>Mixte primaire- secondaire</b>
Petite	Moins de 200 élèves	Moins de 300 élèves	Moins de 60 élèves
Moyenne	De 200 à 350 élèves	De 300 à 700 élèves	De 60 à 200 élèves
Grande	Plus de 350 élèves	Plus de 700 élèves	Plus de 200 élèves