

Enquête sur le développement des enfants en contexte de pandémie – volet « parent »

Méthodologie de l'enquête

Institut de la statistique du Québec

Juillet 2022

Publication réalisée
à l'Institut de la statistique du Québec par : Marilène Courteau et Amélie Lavoie

Avec la collaboration de : Camille Thériault-Marois

Sous la coordination de : Marie-Eve Tremblay et Nathalie Audet

Révision linguistique et mise en page : Direction de la diffusion et des communications

Enquête financée par : Ministère de la Santé et des Services Sociaux
Université de Montréal
Ministère de l'Éducation

Enquête sous la responsabilité de : Direction principale des statistiques sociales et de santé
Institut de la statistique du Québec

Pour tout renseignement concernant
le contenu de ce document : Direction de la méthodologie
Institut de la statistique du Québec
200, chemin Sainte-Foy, 3^e étage
Québec (Québec) G1R 5T4
Téléphone : 418 691-2410
ou
1 800 463-4090

Notice bibliographique suggérée pour le rapport

COURTEAU, Marilène, et Amélie LAVOIE (2022). *Enquête sur le développement des enfants en contexte de pandémie – volet « parent »*. *Méthodologie de l'enquête*, [En ligne], Québec, Institut de la statistique du Québec, 21 p.

Avertissement

En raison de l'arrondissement des données, le total ne correspond pas nécessairement à la somme des parties. Afin de faciliter la lecture des résultats, les proportions de 5 % et plus sont arrondies à l'unité dans le corps du texte. Ainsi, les proportions dont la décimale est de ,5 ont été arrondies à l'unité inférieure ou supérieure selon la seconde décimale. Par exemple, la proportion 20,47 % est arrondie à 20 %, alors que la proportion 20,53 % est arrondie à 21 %. Les arrondissements des proportions se terminant par ,5 dans les tableaux et figures varient donc dans le corps du texte.

Note : Le genre masculin a été utilisé dans le but d'alléger le texte.

Table des matières

Introduction.....	6
1.1 Plan d'échantillonnage.....	7
1.1.1 Population visée	7
1.1.2 Base de sondage.....	8
1.1.3 Méthode de sélection, taille et répartition de l'échantillon.....	8
1.2 Collecte de données.....	10
1.2.1 Questionnaire d'enquête	10
1.2.2 Prétest.....	10
1.2.3 Stratégie de collecte.....	10
1.2.4 Résultats de la collecte.....	11
1.3 Traitement des données	12
1.3.1 Validation	12
1.3.2 Pondération.....	13
1.3.3 Non-réponse partielle	15
1.4 Précision des estimations	17
1.5 Portée et limites de l'enquête	18
Conclusion.....	19
Références bibliographiques	20
Annexe.....	21

Introduction

L'Institut de la statistique du Québec (ISQ) s'est fait confier par l'Observatoire pour l'éducation et la santé des enfants (OPES) le mandat de réaliser l'*Enquête sur le développement des enfants en contexte de pandémie* en collaboration avec le ministère de l'Éducation du Québec (MEQ). Créé en 2021, l'OPES cherche à comprendre les répercussions qu'aura eues la pandémie de COVID-19 sur les enfants. Ces derniers peuvent en effet avoir subi des perturbations tant sur le plan scolaire que sur le plan familial, et un portrait juste de la situation aidera les décideurs à prendre les mesures qui s'imposent en matière d'éducation, de développement psychosocial, de santé et de bien-être afin d'atténuer les conséquences négatives de la crise sanitaire sur les jeunes générations d'aujourd'hui. L'OPES s'intéresse à différents thèmes tels la santé mentale, les habitudes de vie, le retard scolaire, le décrochage scolaire, les relations sociales et familiales ainsi que l'utilisation des services sociaux et de santé.

Pour atteindre ses objectifs de recherche, l'OPES utilisera différentes données, dont des données administratives et des données d'enquêtes déjà menées ou qui seront menées dans le futur, et s'inspirera des travaux de l'ISQ et de chercheurs. L'OPES souhaite, entre autres, suivre dans le temps les enfants qui fréquentaient la maternelle 5 ans en 2016-2017 et qui avaient été visés par l'*Enquête sur le développement des enfants à la maternelle 2017* (EQDEM 2017). Grâce aux données recueillies dans le cadre de cette enquête, laquelle a été réalisée quelques années avant la crise sanitaire, il sera possible de mettre en relation l'état de développement de ces enfants tel qu'il était avant la pandémie avec celui d'aujourd'hui.

L'*Enquête sur le développement des enfants en contexte de pandémie* se divise en deux volets : le volet « parent » et le volet « enseignant ». Pour le premier, la collecte de données a été réalisée au cours de l'été 2021 auprès des parents d'enfants pour lesquels un questionnaire avait été rempli dans le cadre de l'EQDEM en 2017. Pour le deuxième, des données ont été recueillies auprès des enseignants et enseignantes de cinquième année du primaire au printemps 2022. Le présent rapport porte sur le premier volet, soit le volet « parent ».

Objectifs de l'enquête

Le volet « parent » de l'enquête vise principalement à faire état de l'expérience vécue par les enfants et leur famille durant la pandémie de COVID-19. Un jumelage des données recueillies dans le cadre de ce premier volet avec celles obtenues dans le cadre du volet « enseignant », ainsi qu'avec celles issues de l'EQDEM 2017, permettra de comprendre comment la pandémie de COVID-19 a pu affecter différents aspects du développement des enfants qui étaient en cinquième année du primaire durant l'année scolaire 2021-2022.

L'enquête vise plus spécifiquement à recueillir de l'information sur les aspects suivants :

- la situation familiale de l'enfant;
- le comportement de l'enfant;

- l'expérience de l'enfant en lien avec la COVID-19 (p. ex. le nombre de tests passés, les absences pendant l'année scolaire en raison de la COVID-19, son expérience de l'école à distance, l'accompagnement fourni et le soutien reçu, les difficultés d'apprentissage et le rendement scolaire, les répercussions de la pandémie sur différents aspects de sa vie, etc.);
- l'expérience des parents en lien avec la COVID-19 (p. ex. les répercussions de la pandémie sur différents aspects de leur vie, comme leur santé mentale).

La première section de ce rapport présente le plan d'échantillonnage de l'enquête (population visée, base de sondage, échantillonnage), alors que la deuxième décrit brièvement les opérations de collecte de données et le bilan de celles-ci. La troisième section fournit une description des différentes étapes du traitement des données. La quatrième section porte sur la précision des données et la présentation des résultats. Enfin, le rapport se conclut par la portée et les limites de l'enquête.

1.1 Plan d'échantillonnage

1.1.1 Population visée

La population visée par le volet « parent » de *l'Enquête sur le développement des enfants en contexte de pandémie* correspond à l'ensemble des enfants qui avaient été visés par l'EQDEM 2017 et qui étaient en quatrième année du primaire dans une école francophone ou anglophone, publique ou privée (subventionnée ou non), du Québec durant l'année scolaire 2020-2021. Précisons que les enfants qui avaient été visés par l'EQDEM 2017 étaient ceux qui fréquentaient la maternelle 5 ans à temps plein en 2016-2017, dans une école francophone ou anglophone, publique ou privée (subventionnée ou non), du Québec; toutefois, les enfants qui résidaient dans les régions sociosanitaires 17 (Nunavik) ou 18 (Terres-Cries-de-la-Baie-James) étaient exclus¹.

Sont donc exclus de la population visée par *l'Enquête sur le développement des enfants en contexte de pandémie* les enfants ayant redoublé une année scolaire, ceux ayant déménagé à l'extérieur du Québec après la maternelle, les enfants de quatrième du primaire qui ne faisaient pas partie de la population visée par l'EQDEM 2017, comme les nouveaux arrivants, et les enfants résidant dans les régions sociosanitaires 17 (Nunavik) ou 18 (Terres-Cries-de-la-Baie-James).

Il est à noter que bien que la population visée soit composée d'enfants, ce sont les parents qui étaient invités à remplir le questionnaire de l'enquête en leur nom.

¹ Pour obtenir plus de renseignements sur la méthodologie utilisée pour l'EQDEM 2017, voir Tremblay et Simard (2018).

1.1.2 Base de sondage

La base de sondage utilisée pour sélectionner l'échantillon de l'enquête a été élaborée à partir des données recueillies dans le cadre de l'EQDEM 2017. Des informations provenant des fichiers du ministère de l'Éducation du Québec ont été ajoutées à la base de sondage après la sélection de l'échantillon. Ces informations ont entre autres été utilisées pour contacter les parents des enfants sélectionnés.

1.1.3 Méthode de sélection, taille et répartition de l'échantillon

L'un des principaux objectifs de l'enquête est de fournir des estimations fiables à la fois pour l'ensemble de la province, pour le territoire de la région métropolitaine de recensement (RMR) de Montréal et pour le regroupement « Autres régions » – c'est-à-dire l'ensemble des régions administratives du Québec situées complètement en dehors de cette RMR, et une partie de Lanaudière, des Laurentides et de la Montérégie –, et ce, tant pour l'ensemble de la population visée que pour la sous-population comptant les 10 % d'enfants les plus vulnérables. Le plan de sondage de cette enquête a donc été construit de manière à ce qu'il soit possible² :

- d'estimer une proportion de 10 % ou plus pour la sous-population comptant les 10 % d'enfants les plus vulnérables avec une bonne précision relative, c'est-à-dire un coefficient de variation (CV)³ d'au plus 15 % à l'échelle provinciale;
- d'estimer une proportion de 10 % ou plus pour la sous-population comptant les 10 % d'enfants les plus vulnérables avec une bonne précision relative, c'est-à-dire un coefficient de variation (CV)⁴ d'au plus 15 % à l'échelle régionale (c'est-à-dire pour la RMR de Montréal et pour le regroupement « Autres régions »).

Le nombre total de répondants nécessaire à l'atteinte des objectifs mentionnés précédemment a été fixé à 8 800 enfants. Cependant, étant donné la durée de la collecte, le moment où elle a été faite (été 2021) ainsi que la disponibilité limitée des ressources pour l'effectuer, la taille d'échantillon pour le volet « parent » a dû être révisée à la baisse. Le nombre maximum de répondants a donc été fixé à 4 116. Mentionnons que nous avons pris soin de garder la répartition entre les strates proportionnelle à la répartition de l'échantillon qui avait été déterminée initialement. Ce nombre de répondants réduit devrait permettre :

- d'estimer une proportion de 11 % ou plus pour la sous-population comptant les 10 % d'enfants les plus vulnérables avec une bonne précision relative, c'est-à-dire un coefficient de variation (CV)⁵ d'au plus 15 % à l'échelle provinciale;

² Dans la mesure où les hypothèses de collecte (taux de réponse et taux d'admissibilité) sont valides.

³ Des renseignements sur le calcul du coefficient de variation sont fournis à la section 1.4.

⁴ Des renseignements sur le calcul du coefficient de variation sont fournis à la section 1.4.

⁵ Des renseignements sur le calcul du coefficient de variation sont fournis à la section 1.4.

- d'estimer une proportion de 20 % ou plus pour la sous-population comptant les 10 % d'enfants les plus vulnérables avec une bonne précision relative, c'est-à-dire un coefficient de variation (CV)⁶ d'au plus 15 % à l'échelle régionale (c'est-à-dire pour la RMR de Montréal et pour le regroupement « Autres régions »).

Ainsi, finalement, un échantillon de 9 083 enfants a été sélectionné pour l'enquête. En faisant des hypothèses quant aux taux de réponse et d'admissibilité⁷, on a pu déterminer la taille d'échantillon nécessaire pour chacune des strates. Ces strates, visant à assurer un certain contrôle de la composition de l'échantillon, ont été formées par le croisement des variables suivantes : le découpage géographique principal (qui se divise en deux : la RMR de Montréal et le regroupement « Autres régions ») et les régions administratives se trouvant à l'intérieur du découpage géographique principal. Au total, 20 strates ont été formées. Le 0 présente la taille de l'échantillon et le nombre de répondants visés par strate.

Tableau 1.1. Taille de l'échantillon et nombre de répondants visés pour le volet « parent » de l'Enquête sur le développement des enfants en contexte de pandémie

Découpage géographique principal	Région administrative	Taille d'échantillon	Répondants visés (n)
RMR de Montréal	Montréal (06)	1947	892
	Laval (13)	456	216
	Lanaudière (14)	393	185
	Laurentides (15)	438	206
	Montréal (16)	1 209	558
Autres régions	Bas-Saint-Laurent (01)	210	96
	Saguenay–Lac-Saint-Jean (02)	293	138
	Capitale-Nationale (03)	848	378
	Mauricie (04)	256	116
	Estrie (05)	371	160
	Outaouais (07)	515	217
	Abitibi-Témiscamingue (08)	177	81
	Côte-Nord (09)	93	42
	Nord-du-Québec (10)	20	9
	Gaspésie–Îles-de-la-Madeleine (11)	74	34
	Chaudière-Appalaches (12)	539	240
	Lanaudière (14)	188	82
	Laurentides (15)	177	82
	Montréal (16)	600	258
	Centre-du-Québec (17)	279	126
Ensemble de la population visée		9 083	4 116

Source : Institut de la statistique du Québec, volet « parent » de l'Enquête sur le développement des enfants en contexte de pandémie.

⁶ Des renseignements sur le calcul du coefficient de variation sont fournis à la section 1.4.

⁷ Les hypothèses pour le taux d'admissibilité et pour le taux de réponse à l'échelle provinciale sont respectivement de 92 % et 51 %.

1.2 Collecte de données

1.2.1 Questionnaire d'enquête

Le questionnaire de l'enquête visait à recueillir des renseignements sur différents aspects de la vie des enfants dans le contexte de la pandémie et de l'imposition de mesures de confinement par les gouvernements.

En moyenne, les répondants ont pris environ 20 minutes pour remplir le questionnaire sur le Web, tandis qu'il leur a fallu en moyenne 36 minutes pour y répondre par téléphone. Le questionnaire comportait les grandes sections suivantes :

- **Section 1** : Situation familiale
- **Section 2** : Comportement
- **Section 3** : L'expérience COVID-19 de l'enfant
- **Section 4** : Votre expérience de la COVID-19
- **Section 5** : Bien-être
- **Section 6** : Caractéristiques sociodémographiques

1.2.2 Prétest

Aucun prétest n'a été réalisé pour ce volet de l'enquête.

1.2.3 Stratégie de collecte

La collecte des données s'est déroulée du 14 juillet au 14 septembre 2021, soit sur une période de neuf semaines.

Une lettre de présentation de l'enquête a d'abord été envoyée aux parents des enfants sélectionnés pour leur expliquer les objectifs de l'enquête et les inviter à se rendre sur le site Web pour remplir le questionnaire électronique. Environ 10 jours après l'envoi de cette lettre, les parents des enfants échantillonnés pour lesquels le questionnaire électronique n'était pas déjà rempli ont reçu une lettre de relance qui mentionnait l'importance de l'étude et les invitait de nouveau à aller remplir le questionnaire en ligne.

Environ une semaine après l'envoi de la lettre de relance, le personnel de l'ISQ a téléphoné aux parents qui n'avaient pas encore rempli le questionnaire afin de les inciter à le faire sur le Web ou à réaliser l'entrevue téléphonique, selon leur préférence.

La collecte des données s'est effectuée par interview téléphonique assistée par ordinateur (ITAO) et/ou par auto-interview assistée par ordinateur (AIAO) à partir du logiciel Interviewer de Voxco. L'ITAO et l'AIAO présentent des avantages quant à la qualité des données recueillies par rapport aux autres méthodes de collecte. Premièrement, le libellé des questions, comprenant les périodes de référence et les pronoms, est personnalisé automatiquement en

fonction de facteurs, comme l'âge et le sexe du répondant, et des réponses aux questions précédentes. Deuxièmement, des mesures de contrôle isolent les réponses incohérentes ou hors-norme, et des instructions apparaissent à l'écran lorsqu'une telle situation se présente. Le répondant reçoit une rétroaction immédiate, et une correction de l'incohérence peut être apportée. Troisièmement, avec les filtres programmés dans le questionnaire, il est possible de sauter automatiquement les questions qui ne concernent pas le répondant.

On estimait au départ qu'il faudrait 20 minutes pour remplir le questionnaire. Cette estimation ne s'est avérée exacte que pour les questionnaires remplis sur le Web. En effet, dès les premières entrevues téléphoniques, les intervieweurs ont remarqué qu'elles prenaient plus que 20 minutes. Nous avons donc revu notre estimation à 35 minutes dès la première journée d'appel (29 juillet 2021) et avons modifié l'information sur la durée attendue dans le questionnaire téléphonique.

Il est à noter que les répondants ont utilisé le Web pour remplir et soumettre leur questionnaire en forte proportion (environ 90 %).

1.2.4 Résultats de la collecte

Comme les taux de réponse généralement présentés par l'ISQ, ceux présentés pour l'enquête sont pondérés. Ils tiennent compte, entre autres facteurs, de la non-proportionnalité régionale⁸ de l'échantillon de l'enquête.

Taux d'admissibilité

Le taux d'admissibilité pondéré⁹ à l'enquête est de 90,7 %. Parmi les enfants non admissibles (9,3 %), 6,6 % n'étaient pas en quatrième année du primaire (il est possible qu'ils aient redoublé ou sauté une année scolaire) et 2,1 % étaient scolarisés en dehors du Québec. Les autres causes d'inadmissibilité étaient le fait de résider dans les régions sociosanitaires 17 ou 18 et le fait d'être scolarisé à la maison.

Taux de réponse

Au total, 4 524 personnes ont participé à l'enquête, ce qui correspond à un taux de réponse pondéré¹⁰ de 54,7 %.

⁸ Étant donné les objectifs de précision pour le découpage géographique principal (qui se divise en deux : la RMR de Montréal et le regroupement « Autres régions »), la proportion de la taille de l'échantillon par rapport à la taille de la population diffère d'une division à l'autre. Les régions situées en dehors de la RMR de Montréal sont légèrement surreprésentées dans l'échantillon par rapport à leur poids dans la population.

⁹ Taux d'admissibilité pondéré (avec le poids de sélection). Plus de renseignements sur la pondération sont fournis à la section 1.3.2.

¹⁰ Taux de réponse pondéré avec le poids initial ajusté pour l'admissibilité (ajustement décrit à la section 1.3.2).

L'absence de numéros de téléphone ou la présence de mauvais numéros dans la base de sondage ainsi que les refus expliquent un peu moins du tiers de la non-réponse. La non-réponse est attribuable à ces deux raisons dans des proportions respectives de 16 % et 14 %.

Le taux de réponse pondéré est presque identique pour les deux divisions du découpage géographique principal. Il est de 54,8 % pour la RMR de Montréal et de 54,6 % pour le regroupement « Autres régions ». Le tableau suivant présente le nombre de répondants et le taux de réponse pondéré pour chacune d'elle, et pour l'ensemble de la population visée.

Tableau 1.2. Nombre de répondants et taux de réponse pondéré pour chacune des deux divisions du découpage géographique principal et pour l'ensemble de la population visée

Découpage géographique principal	Répondants (n)	Taux de réponse pondéré (%)
RMR de Montréal	2 272	54,8
Autres régions	2 252	54,6
Ensemble de la population visée	4 524	54,7

Source : Institut de la statistique du Québec, volet « parent » de l'*Enquête sur le développement des enfants en contexte de pandémie*.

1.3 Traitement des données

1.3.1 Validation

Le recours aux questionnaires informatisés (ITAO et AIAO) permet une saisie automatique des réponses. La très grande majorité des questions sont autocodées, c'est-à-dire que l'intervieweur ou le répondant n'a qu'à cocher le choix de réponse : la saisie s'en trouve donc réduite au minimum. De plus, comme les questionnaires Web et téléphoniques sont entièrement informatisés, plusieurs validations de base sont réalisées pendant que le répondant répond aux questions sur le Web ou par téléphone, notamment la vérification du respect des choix de réponse pour les questions fermées ou celle de l'adéquation des sauts de section, pour les blocs de questions ne s'adressant qu'à un sous-groupe de répondants.

Des vérifications concernant le statut de réponse des individus (répondants, non répondants ou inadmissibles) ont aussi été effectuées *a posteriori*. Celles-ci nous ont principalement permis de conserver comme répondants certains individus qui n'avaient pas rempli le questionnaire en entier et d'en exclure d'autres qui n'avaient pas répondu à une proportion importante de questions.

De plus, le fichier des données de l'enquête a fait l'objet de vérifications. La première avait pour but de s'assurer que, pour chaque question, tous les répondants qui devaient y répondre l'avaient fait et que tous ceux qui ne devaient pas y répondre ne l'avaient pas fait.

Enfin, on a effectué la vérification des codes postaux fournis par les répondants afin de s'assurer que chaque répondant était classé dans la bonne région administrative et indicateur qui serait créé, tel que l'indice de défavorisation matérielle et sociale, allait être adéquat.

1.3.2 Pondération

La pondération consiste à attribuer à chaque répondant de l'enquête une valeur, c'est-à-dire un poids qui correspond au nombre de personnes qu'il « représente » dans la population. La pondération est essentielle pour la production de diverses estimations à partir des données recueillies dans le cadre de l'enquête. Elle nous permet de rapporter les données des répondants à la population visée et, ainsi, de faire des inférences adéquates à cette population, bien que celle-ci n'ait pas été observée dans sa totalité.

En raison des objectifs de diffusion de l'enquête selon le découpage géographique principal (la RMR de Montréal et le regroupement « Autres régions »), l'échantillon a non seulement été réparti selon ce découpage, mais aussi selon les régions administratives se trouvant à l'intérieur de celui-ci. Toutefois, cette répartition est légèrement différente de celle de la population du Québec. La pondération permet la correction de la distorsion créée par le plan de sondage utilisé pour la sélection de l'échantillon.

De plus, on sait que, dans les enquêtes, la probabilité de répondre varie selon plusieurs caractéristiques sociodémographiques. Il est donc nécessaire de tenir compte de ces éléments en les intégrant à la pondération, qui sera utilisée lors de la production de statistiques à partir des données de l'enquête. La pondération est également l'un des éléments à prendre en compte pour estimer correctement la précision des données.

La stratégie de pondération établie pour cette enquête tient compte, entre autres, de la probabilité qu'une personne soit sélectionnée dans l'échantillon et de la portion de l'échantillon qui s'est avérée inadmissible lors de la collecte des données. Elle comprend également un ajustement pour la non-réponse totale à l'enquête.

Description des étapes

La première étape consiste à calculer le poids initial de chaque individu faisant partie de l'échantillon. Étant donné que la base de sondage est constituée des enfants pour lesquels un questionnaire a été rempli dans le cadre de l'EQDEM 2017, le poids initial d'un individu est égal au produit obtenu lorsqu'on multiplie le poids qu'il avait dans le cadre de l'EQDEM 2017 (« poids EQDEM 2017 ») par un facteur d'ajustement. Ce facteur d'ajustement est le même pour tous les individus d'une même strate. En effet, pour chacune des strates, la somme des poids initiaux des individus de l'échantillon doit être égale à la somme des poids EQDEM 2017 de tous les individus de la strate. Ainsi, le facteur d'ajustement est égal au quotient que l'on obtient lorsqu'on divise la somme des poids EQDEM 2017 de tous les individus de la strate par la somme des poids EQDEM 2017 des individus de la strate faisant partie de l'échantillon.

Ensuite, on doit ajuster le poids initial de chaque individu pour tenir compte de l'inadmissibilité de certaines personnes à l'enquête. Cet ajustement se fait en deux étapes : on doit d'abord le faire en fonction de l'inadmissibilité établie préalablement à la collecte et ensuite selon l'inadmissibilité établie en cours de collecte. En effet, pour cette enquête, les données obtenues du MEQ après la sélection de l'échantillon permettent d'identifier a priori certains enfants non admissibles avant même la collecte de données auprès des parents (enfant n'étant pas en quatrième année du primaire, enfant scolarisé en dehors du Québec, enfant résidant dans les régions sociosanitaires 17 ou 18). Pour ces enfants, considérés comme inadmissibles avant même la collecte, le poids est égal à zéro. Par la suite, pour les enfants restants, le statut d'admissibilité final étant inconnu pour plusieurs non-répondants, il faut réduire leur poids pour tenir compte du fait qu'une partie d'entre eux est probablement inadmissible. Pour ce faire, on multiplie le poids de ces enfants par le taux d'admissibilité établie en cours de collecte pondéré, qui correspond à la proportion pondérée (avec le poids initial) d'enfants admissibles parmi ceux pour lesquels l'admissibilité a pu être déterminée lors de l'enquête.

L'étape suivante est l'ajustement pour la non-réponse totale. À cet effet, on effectue une analyse pour définir les variables ou caractéristiques selon lesquelles la probabilité d'avoir un questionnaire considéré comme rempli pourrait varier. On utilise la méthode du score¹¹ pour déterminer quels sont les groupes homogènes de pondération, c'est-à-dire des sous-groupes de l'échantillon à l'intérieur desquels la propension à répondre est assez uniforme¹². L'ajustement s'effectue ensuite à l'intérieur de chacun de ces groupes : le poids des enfants pour lesquels un questionnaire considéré comme rempli est disponible est ajusté par l'inverse du taux de réponse (pondéré) dans le groupe homogène de pondération.

Par la suite, on vérifie qu'aucun enfant n'a de poids très élevé comparativement au poids des enfants d'un même découpage géographique ou d'une même région administrative. Seuls deux cas de figure ont été relevés dans le cadre de cette enquête. Dans ces cas, le poids sera égal au poids maximal de la strate (formée par le croisement du découpage géographique et de la région) dont ils font partie en excluant les poids extrêmes, puis les poids de la strate sont ensuite ajustés selon un même facteur (facteur = poids total avant correction/poids total après correction) pour que le poids total de la strate après correction des poids extrêmes soit le même qu'avant la correction.

¹¹ Pour obtenir plus de renseignements sur la méthode du score, voir Haziza et Beaumont (2007) et Eltinge et Yansaneh (1997).

¹² Les variables (qui sont créées à partir des données obtenues du MEQ ou des données tirées de l'EQDEM 2017, ou qui sont issues de ces deux sources) retenues dans le modèle sont : la région, la langue maternelle, l'indice de défavorisation matérielle et sociale et la présence d'une vulnérabilité dans au moins deux domaines de développement selon l'EQDEM 2017.

1.3.3 Non-réponse partielle

Définition

La pondération tient compte de la non-réponse totale, mais n'a pas été ajustée en fonction de la non-réponse partielle. Cette dernière se produit lorsqu'une variable (ou lorsqu'un indicateur créé à partir de plusieurs questions) comporte des données manquantes pour les répondants à l'enquête. On sait qu'une non-réponse partielle importante peut entraîner certains biais dans les estimations, au même titre que la non-réponse totale, s'il s'avère que les non-répondants présentent des caractéristiques différentes de celles des répondants et, de surcroît, si ces caractéristiques sont liées au thème abordé. Le taux de non-réponse partielle pondéré pour une variable donnée se définit comme le rapport entre le nombre pondéré de personnes qui n'ont pas répondu à la question et le nombre pondéré de personnes qui auraient dû y répondre. Plus ce taux est élevé, plus le risque de biais induits par la non-réponse partielle est grand. En général, lorsque l'on analyse des données d'enquête, il est raisonnable de faire l'hypothèse qu'un taux de non-réponse partielle inférieur à 5 % a un effet négligeable sur les estimations à l'échelle provinciale, sauf dans la situation où cette non-réponse est très caractérisée. Les variables pour lesquelles le taux de non-réponse partielle dépasse ce seuil de 5 % doivent en général être analysées plus en profondeur au regard des biais. En raison des faibles taux de non-réponse partielle observés pour la grande majorité des variables de l'enquête, les risques de biais, relativement à la précision des estimations, sont assez faibles. Les variables ayant un taux de non-réponse partielle supérieur à 5 % sont présentées dans le *Tableau A1 – Variables ayant un taux de non-réponse partielle pondéré supérieur à 5 %, Québec, 2022* se trouvant en annexe.

Analyse de la non-réponse partielle

Une seule variable ayant un taux de non-réponse partielle élevé, soit le revenu du ménage, a fait l'objet d'une imputation afin que le risque de biais soit réduit. Cette imputation est expliquée à la section qui suit.

Les trois autres variables ayant un taux de non-réponse partielle supérieur à 5 % n'ont pas fait l'objet d'une imputation. En effet, ces variables étant de moindre intérêt analytique (dont deux ne portent que sur des sous-populations), l'effet de la non-réponse partielle sur les analyses est jugé négligeable.

Imputation

Revenu du ménage

Le revenu du ménage est une donnée généralement difficile à obtenir des répondants dans une enquête populationnelle, et ce fut effectivement le cas dans le cadre du volet « parent » de l'*Enquête sur le développement des enfants en contexte de pandémie*. Les questions CSD_Q14 et CSD_Q15 à CSD_Q15F portaient sur l'estimation, pour

les 12 derniers mois, du revenu total de l'ensemble des membres du ménage provenant de toutes les sources, avant impôt et autres retenues. Le répondant avait la possibilité de fournir soit une valeur exacte, bien qu'approximative, soit une estimation sous forme de catégorie en indiquant la tranche de revenu¹³ correspondant à la situation de son ménage. Parmi les 4 524 répondants, une proportion pondérée de 2,9 % a refusé ou n'était pas en mesure d'indiquer le revenu du ménage, alors qu'une proportion pondérée de 5,2 % a déclaré le revenu du ménage à l'aide d'une catégorie. Si l'on considère que la déclaration sous forme de catégorie est associée à une non-réponse partielle, on obtient un taux global pondéré de non-réponse partielle de 8,1 %. En présence d'une non-réponse partielle aussi élevée dans les enquêtes, il est recommandé d'imputer les données, et ce, afin de réduire au minimum les risques de biais. Une imputation du revenu total du ménage pour les répondants n'ayant pas fourni de réponse à cette question a donc été faite. Une valeur exacte de revenu a également été imputée pour les répondants qui avaient fourni un revenu sous forme de catégorie, car cette valeur était nécessaire dans plusieurs cas lors de la création d'un indicateur du niveau de revenu du ménage. Pour les répondants qui n'avaient fourni aucune information concernant le revenu de leur ménage, le revenu exact a été imputé par régression linéaire¹⁴. Pour les répondants qui avaient fourni le revenu uniquement sous forme de catégorie, le revenu exact a été imputé séparément par catégorie de revenu à l'aide de l'imputation aléatoire par donneur (*Hot deck*) à l'intérieur de classes d'imputation. Pour chaque catégorie de revenu, on a d'abord créé les classes d'imputation par la méthode du score avec des caractéristiques du ménage et du répondant liées significativement au revenu du ménage¹⁵. L'imputation de données a pour effet d'augmenter artificiellement le nombre de répondants ayant une donnée de revenu du ménage, sans que l'on en tienne toutefois compte dans l'estimation de la variance selon les méthodes habituelles, ce qui fait que la variance des estimations faisant intervenir le niveau de revenu est sous-estimée. Il est important de tenir compte de cette réalité lors de l'interprétation des résultats. Par

¹³ Les catégories de revenu proposées étaient les suivantes : moins de 10 000 \$, de 10 000 \$ à moins de 20 000 \$, de 20 000 \$ à moins de 30 000 \$, de 30 000 \$ à moins de 40 000 \$, de 40 000 \$ à moins de 50 000 \$, de 50 000 \$ à moins de 60 000 \$, de 60 000 \$ à moins de 70 000 \$, de 70 000 \$ à moins de 80 000 \$, de 80 000 \$ à moins de 100 000 \$, de 100 000 \$ à moins de 120 000 \$ et 120 000 \$ et plus. De plus, pour certaines catégories de revenu, une question supplémentaire était posée au répondant. Pour les catégories « de 10 000 \$ à moins de 20 000 \$ », « de 20 000 \$ à moins de 30 000 \$ », « de 30 000 \$ à moins de 40 000 \$ », « de 40 000 \$ à moins de 50 000 \$ », « de 50 000 \$ à moins de 60 000 \$ » et « de 60 000 \$ à moins de 70 000 \$ », on demandait si le revenu se situait dans la tranche de 5 000 \$ supérieure de la catégorie.

¹⁴ Deux modèles différents ont été construits : un pour les répondants dont au moins un des conjoints travaille et un pour les répondants dont aucun des conjoints ne travaille. Les variables explicatives retenues dans au moins l'un des deux modèles sont : le fait que l'enfant soit scolarisé dans le réseau public ou privé, la langue d'usage, le lieu de naissance (au moins un des conjoints est né au Canada / les conjoints sont nés ailleurs qu'au Canada), l'indice de défavorisation matérielle et sociale, le nombre de personnes dans le ménage (2, 3, 4 et plus), le fait d'habiter avec un conjoint, le niveau de scolarité le plus élevé des conjoints, le nombre de semaines travaillées (moins de 39 semaines / 39 semaines et plus) par le conjoint ayant travaillé le plus de semaines et le nombre d'heures travaillées par semaine par ce même conjoint..

¹⁵ Les variables ayant servi à construire les classes d'imputation sont la région, la langue d'usage, le fait d'habiter avec un conjoint, le niveau de scolarité le plus élevé des conjoints, le nombre de semaines (moins de 39 semaines / 39 semaines et plus) travaillées par le conjoint ayant travaillé le plus de semaines et le nombre d'heures travaillées par semaine par ce même conjoint.

exemple, il est recommandé d'interpréter avec prudence les différences significatives dont le seuil observé se trouve très près du seuil de signification.

1.4 Précision des estimations

Dans toute enquête statistique, les estimations produites sont entachées d'erreurs. Il y a d'une part l'erreur liée à l'échantillonnage (voir section suivante), qui est attribuable au fait que seule une partie de la population visée est sélectionnée pour participer à l'enquête, et, d'autre part, les erreurs qu'entraînent, par exemple, la non-réponse et les erreurs de mesure associées à des questions difficiles à interpréter, à la saisie ou au traitement de données. Il a déjà été question, dans les sections précédentes, de la principale source d'erreur non liée à l'échantillonnage, soit la non-réponse.

L'erreur due à l'échantillonnage

Ce type d'erreur se répercute sur les estimations produites, dont la précision est influencée par la complexité du plan d'échantillonnage. Il est donc nécessaire de mesurer la précision de chaque estimation avant de procéder à l'analyse et à l'interprétation des résultats inférés à la population visée. Pour cette enquête, une série de 500 poids d'autoamorçage, aussi appelés « poids *bootstrap* », a été créée afin de tenir compte adéquatement non seulement du plan de sondage complexe, mais aussi de tous les ajustements pour l'inadmissibilité et la non-réponse apportés à la pondération dans l'estimation de la variance et les tests statistiques.

Le coefficient de variation (CV), qui permet la mesure de la précision relative d'une estimation, est utilisé pour quantifier l'erreur d'échantillonnage. Plus le CV est petit, plus l'estimation est précise. Le CV s'exprime comme suit :

$$CV = \frac{\sqrt{\text{variance de l'estimation}}}{\text{estimation}}$$

Une autre mesure de la précision d'une estimation est l'étendue de l'intervalle de confiance (IC) : plus l'intervalle est court, plus la valeur du paramètre est circonscrite. L'intervalle de confiance associé à une proportion représente la zone d'incertitude liée à l'estimation. Cette zone d'incertitude est étroitement liée au niveau de confiance choisi. Ainsi, un intervalle de confiance pour une proportion à un niveau de confiance de 95 % signifie que si l'on répétait l'enquête 100 fois et qu'à chaque fois on estimait la proportion et calculait son intervalle de confiance, 95 des 100 intervalles ainsi créés contiendraient la vraie valeur de la proportion dans la population.

Présentation des résultats

Lors de la présentation de résultats tirés de l'enquête, il est important de rendre compte du fait que les statistiques fournies sont des estimations et non des valeurs exactes.

1.5 Portée et limites de l'enquête

La participation des individus à l'enquête a été bonne, comme le démontre le taux de réponse obtenu lors de la collecte des données (54,7 %), qui est supérieur à celui prévu (51 %). Le potentiel analytique des données de l'enquête se rapportant à l'ensemble des répondants est bon, et la précision obtenue est celle visée au moment de la conception du plan de sondage.

La pondération a permis de réduire les biais engendrés par la non-réponse dans les estimations, notamment grâce à des variables explicatives présentes dans la base de sondage.

Par ailleurs, comme c'est le cas dans la plupart d'enquêtes populationnelles comportant des données autodéclarées, il est difficile de garantir l'exactitude des réponses fournies par les répondants. Ces derniers peuvent être influencés par le phénomène de désirabilité sociale ou par la difficulté de se rappeler des choses du passé.

De plus, dans le cas où les résultats de l'enquête permettent de déceler des liens entre deux variables de même que des différences entre des groupes de population, il est important de noter qu'ils ne permettent pas d'établir de lien de causalité entre les caractéristiques étudiées. Mentionnons qu'il convient également de tenir compte de facteurs confondants dans l'analyse des liens entre variables.

Finalement, il est important de tenir compte des mises en garde et des recommandations mentionnées dans les sections précédentes afin d'utiliser adéquatement les données de cette enquête. La section 1.4 explique, entre autres, comment tenir compte du plan de sondage complexe lors de la mesure de la précision des résultats produits. Il est à noter qu'une analyse de la non-réponse partielle devrait être effectuée pour tous les résultats qui seront produits afin de vérifier si cette non-réponse est importante et problématique.

Conclusion

Le présent document contient les informations méthodologiques relatives au volet « parent » de *l'Enquête sur le développement des enfants en contexte de pandémie*. La population visée y est présentée. Des informations sur le plan d'échantillonnage et la collecte de données y sont également incluses afin que les conditions et le contexte dans lequel l'enquête a été réalisée soient bien situés. Comme l'enquête a été faite de manière à ce que ses résultats soient inférés à la population visée, une section portant sur le traitement des données, notamment les méthodes de pondération appliquées, se trouve également dans ce rapport. Enfin, des informations relatives aux méthodes d'analyse des données issues de l'enquête sont fournies.

On retiendra que l'enquête a permis de joindre, entre le 14 juillet et le 14 septembre 2021, près de 4 500 parents d'enfants qui avaient été visés par l'EQDEM 2017 et qui étaient en quatrième année du primaire dans un établissement du Québec durant l'année scolaire 2020-2021. La participation à l'enquête a été bonne, comme en témoigne le taux de réponse de 55 %, qui est supérieur à celui qui était attendu. Le potentiel analytique de l'enquête est bon, surtout pour les estimations visant l'ensemble de la population. Mentionnons aussi que très peu de questions sont associées à un taux de non-réponse partielle supérieur à 5 %.

L'ensemble des informations présentées dans ce rapport vise à guider l'utilisateur des données dans son analyse et son interprétation des résultats du volet « parent » de l'enquête.

Références bibliographiques

ELTINGE, J. L., et I. S. YANSANEH (1997). « Méthodes diagnostiques pour la construction de cellules de correction pour la non-réponse, avec application à la non-réponse aux questions sur le revenu de la U.S. Consumer Expenditure Survey », *Techniques d'enquête*, [En ligne], produit n° 12-001-X19970013103 au catalogue de Statistique Canada, vol. 23, n° 1, juin, p. 37-45. [www150.statcan.gc.ca/n1/fr/pub/12-001-x/1997001/article/3103-fra.pdf?st=YRiuivNK].

GAMACHE, P., D. HAMEL et R. PAMPALON (2017). *L'indice de défavorisation matérielle et sociale : en bref*, [En ligne], Institut national de santé publique du Québec, p. 9 [www.inspq.qc.ca/sites/default/files/santescope/indice-defavorisation/guidemethodologiquefr.pdf] (Consulté le 16 mai 2019).

HAZIZA, D., et J.-F. BEAUMONT (2007). "On the Construction of Imputation Classes in Surveys", *International Statistical Review*, [En ligne], vol. 75, n° 1, avril, p. 25-43. [www.jstor.org/stable/41508447].

INSTITUT DE LA STATISTIQUE DU QUÉBEC en collaboration avec l'INSTITUT NATIONAL DE SANTÉ PUBLIQUE DU QUÉBEC (2016). *Guide spécifique des aspects méthodologiques des données d'enquêtes sociosanitaires du Plan national de surveillance de l'état de santé de la population et de ses déterminants (Plan commun de surveillance et Plan ministériel de surveillance multithématique). Enquête québécoise sur la santé de la population 2014-2015*, Québec, Gouvernement du Québec, 83 p.

KORN, E. L., et B. I. GRAUBARD (1999). *Analysis of Health Surveys*, New York, John Wiley & Sons, 382 p.

RUST, K. F., et J. N. K. RAO. (1996). "Variance estimation for complex surveys using replication techniques", *Statistical Methods in Medical Research*, [En ligne], vol. 5, n° 3, septembre, p. 283-310. doi : [10.1177/096228029600500305](https://doi.org/10.1177/096228029600500305).

TREMBLAY, M.-E., et M. SIMARD (2018). *Méthodologie de l'Enquête québécoise sur le développement des enfants à la maternelle 2017*, [En ligne], Québec, Institut de la statistique du Québec, 40 p. [statistique.quebec.ca/fr/fichier/methodologie-de-lenquete-quebecoise-sur-le-developpement-des-enfants-a-la-maternelle-2017.pdf]

Annexe

Tableau A.1 Variables ayant un taux de non-réponse partielle pondéré supérieur à 5 %, Québec, 2022

Nom de la variable	Libellé	Taux de non-réponse partielle pondéré	Effet possible sur l'analyse
SF_Q3	Est-ce que votre conjoint ou conjointe est l'autre parent (biologique ou adoptif) de (Prénom enfant)?	11,8 %	Négligeable
COV_Q3	Au cours de l'année scolaire 2020-2021, environ combien de temps a-t-il/elle passé en confinement en raison de cas de COVID-19 au sein de sa classe, de son école ou de sa famille?	6,3 %	Négligeable
COV_Q13	Au début de la crise liée à la COVID-19, soit de la mi-mars 2020 à la fin de l'année scolaire 2019-2020, environ combien de semaines (Prénom enfant) a-t-il/elle été à l'école en personne?	5,6 %	Négligeable
CSD_Q14	Revenu du ménage	8,1 %	Sous-estimation de la variance due à l'imputation effectuée

Source : Institut de la statistique du Québec, volet « parent » de l'*Enquête sur le développement des enfants en contexte de pandémie*.