



N° 1 | 2023
2023

Les enquêtes Génération

Des enquêtes sur le lien entre formation initiale et insertion professionnelle

Stéphane Jugnot

Édition électronique :

URL : <https://demc-journal.org/articles/revue-1/3350-les-enquetes-generation>

DOI : 10.34745/numerev_1954

ISSN : 3036-5295

Date de publication : 14/12/2023

Cette publication est **sous licence CC-BY-NC-ND** (Creative Commons 2.0 - Attribution - Pas d'Utilisation Commerciale - Pas de Modification).

Pour **citer cette publication** : Jugnot, S. (2023). Les enquêtes Génération . *DEMC Journal*, (1).
https://doi.org/10.34745/numerev_1954

Depuis la fin des années 1990, le Centre d'études et de recherches sur les qualifications (CEREQ) produit les enquêtes « Génération ». A intervalle régulier, il interroge un échantillon de jeunes sortis de formation initiale au cours ou à l'issue d'une année scolaire donnée, quel que soit leur niveau de formation. Ils sont interrogés trois ans après leur sortie et parfois réinterrogés ultérieurement sur leurs premiers pas dans la vie active et sur d'autres dimensions pouvant expliquer les inégalités d'accès à l'emploi. Ces enquêtes, qui relèvent de la Statistique publique au sens de la loi de juin 1951, sont les seules à permettre d'étudier le lien entre formation initiale et insertion professionnelle de façon homogène quel que soit le niveau de sortie. Cet article présente les objectifs de ces enquêtes, leur contenu et les grandes étapes de leur production. Il en propose une vision synoptique car il existe des variations de champ, de thématiques et de dimensionnement de l'échantillon d'une cohorte à l'autre.

Mots-clefs :

Insertion, Inégalités, Jeunes, Panel, Marché du travail

Introduction

Cet article propose une présentation synoptique des enquêtes « Génération » du CEREQ à partir de la « Génération 1998 ». Il en présente les objectifs et leurs intérêts par rapport à d'autres dispositifs d'observation de l'insertion des sortants de formation initiale. Une présentation de la structure du questionnaire permet de préciser la nature des informations proposées. L'article aborde aussi la production des enquêtes, du tirage de l'échantillon à la construction des bases mises à disposition des chercheurs. Il présente les principales évolutions intervenues d'une enquête à l'autre, parfois susceptibles de compliquer les analyses comparatives. Au fil de l'article, il propose une série de références bibliographiques, certaines illustrant les utilisations possibles des données, d'autres relevant de la littérature grise pour des présentations méthodologiques plus détaillées.

Le Centre d'études et de recherches sur les qualifications (CEREQ) est un établissement public chargé d'éclairer les liens entre la formation, les qualifications, l'emploi et le travail, à partir d'études statistiques et d'analyses sociologiques plus qualitatives, conformément à ses missions définies par l'article D313-38 du code de l'Education. Il est placé sous la tutelle du ministère chargé de l'éducation nationale et du ministère

chargé du travail.

Jusqu'à la fin des années 1980, le CEREQ a réalisé des enquêtes statistiques segmentées sur le cheminement des jeunes à l'issue de leur formation initiale, sur les sortants de CAP ou BEP, sur des sortants de l'enseignement secondaire général ou technique, etc.

Dans les années 1990, le CEREQ a décidé de mettre en place un dispositif d'observation homogène, couvrant l'ensemble des sortants de formation initiale, quel que soit leur niveau de sortie. Une première cohorte, ou « Génération », a été enquêtée en 1997. Elle porte sur les jeunes sortis du système éducatif au cours ou à l'issue de l'année scolaire 1991-1992. C'est la « Génération 1992 » mais 20% des sortants de formation initiale échappent encore à cette enquête. L'enquête suivante, collectée en 2001 auprès de la « Génération 1998 » couvre l'ensemble des sortants (Lopez, Epiphane et al., 2002). Elle inaugure un système régulier d'enquêtes : une nouvelle cohorte est interrogée tous les trois ans, trois ans après leur sortie du système éducatif, avec une alternance de cohortes « pleines », réinterrogées cinq ans, sept ans, voire dix ans, après leur sortie et les cohortes « légères », qui ne sont pas réinterrogées. La « Génération 2017 » initie un nouveau rythme^[1] (les enquêtes postérieures à la première interrogation de la « Génération 2017 » ne sont pas présentées dans cet article).

Les enquêtes « Génération » sont labellisées « Statistique publique » au sens de la loi du 7 juin 1951 sur l'obligation, la coordination et le secret en matière de statistiques. Elles sont les seules à permettre d'étudier le lien entre la formation initiale et l'insertion professionnelle de façon homogène quel que soit le niveau et la formation de sortie, du jeune sorti du secondaire sans aucune diplôme jusqu'au titulaire d'un doctorat.

Les enquête « Génération » sont principalement mobilisées pour étudier les conditions différenciées d'accès à l'emploi selon le niveau de sortie, selon le plus haut diplôme atteint et selon la nature de la formation, générale ou professionnelle, par apprentissage ou non (Le Bayon, Dabet et al., 2022). Les caractéristiques socio-démographiques collectées permettent également d'analyser les inégalités selon le genre (Couppié, Epiphane, 2019), l'origine sociale (Vallet, Goffette, 2022), l'origine migratoire (Jugnot, 2023) ou le lieu de résidence (Couppié, Dieusaert et al., 2020).

Principalement construites pour des analyses nationales, certaines enquêtes permettent de produire des résultats régionaux, notamment pour les territoires ayant financé des extensions d'échantillon ad hoc^[2]. Certaines enquêtes approfondissent des thèmes liés aux objectifs du dispositif pour répondre à des demandes d'organismes ou d'institutions partenaires.

1. Présentation du dispositif « Génération »

1.1-Les objectifs des enquêtes « Génération »

Les enquêtes « Génération » sont construites pour étudier la relation entre la formation

initiale et les conditions d'insertion professionnelle des jeunes, quel que soit leur niveau de sortie, qu'ils sortent diplômés ou non. A intervalle régulier, elles interrogent un échantillon de jeunes sur leur parcours scolaire et sur leurs premiers pas dans la vie active.

La connaissance du parcours scolaire permet de ne pas se limiter au diplôme de sortie, à la différence des nombreuses enquêtes d'insertion produites dans une optique d'évaluation de la performance des formations. L'enquête identifie le plus haut diplôme atteint et s'intéresse aux moments clefs de l'orientation, après la troisième et après le baccalauréat. Les expériences de travail en cours de scolarité sont aussi interrogées.

De son côté, l'insertion est abordée dans une perspective multidimensionnelle qui dépasse la simple situation d'activité. La nature de l'emploi occupé et les conditions d'emploi sont abordées, de même que des dimensions subjectives, comme la satisfaction dans l'emploi, les perspectives professionnelles ou le sentiment de discrimination. L'ensemble du parcours suivi au cours des premières années de vie active est caractérisé de façon à observer la vitesse et les étapes de l'accès à l'emploi durable.

Le questionnement permet aux analyses de tenir compte de certaines caractéristiques sociodémographiques des jeunes dès lors qu'elles sont susceptibles d'influencer les conditions d'accès à l'emploi et les emplois occupés, notamment la décohabitation, le genre, les origines sociales, l'ascendance migratoire ou le lieu de résidence.

Tous les jeunes d'une même « Génération » sortent du système éducatif au cours ou à l'issue de la même année scolaire. Ils entrent donc sur le marché du travail dans un contexte macroéconomique identique et tous sont enquêtés avec le même protocole et le même questionnaire. De ce fait, les effets des niveaux de sortie, des types de diplôme, de leurs spécialités ou du passage par l'apprentissage peuvent être évalués de façon beaucoup plus simple que si l'on devait recourir à des données issues de dispositifs segmentés et disparates.

Le dispositif s'appuie enfin sur le constat que l'insertion prend du temps. L'accès à l'emploi à durée interminée, qui reste une norme d'emploi, est rarement immédiat. Aussi, les répondants sont-ils enquêtés avec du recul. De la « Génération 1998 » à la « Génération 2013 », une nouvelle cohorte de sortants est interrogée tous les trois ans, trois ans après leur sortie de formation initiale. Une cohorte sur deux fait l'objet de deux ré-interrogations ultérieures pour disposer d'un parcours sur cinq ans, puis sept ans. La « Génération 1998 » a même fait l'objet d'une dernière interrogation dix ans après la sortie de formation initiale. Cette alternance de cohortes « légères », à l'échantillon plus réduit, et de cohortes « pleines », avec des échantillons plus importants, permet de répondre à la fois à une demande de résultats synthétiques récents et de données permettant des analyses structurelles plus détaillées et sur une plus longue période.

A partir de la « Génération 2017 », le rythme du dispositif est profondément remanié : une nouvelle cohorte doit être interrogée tous les quatre ans au lieu de tous les trois

ans mais toutes les cohortes doivent faire l'objet d'une seconde interrogation au lieu d'une cohorte sur deux. La première interrogation reste placée trois ans après la sortie du système éducatif tandis que la seconde interrogation est repoussée à six ans après la sortie.

Le schéma 1 présente le calendrier de réalisation des enquêtes par « Génération ».

Schéma 1. Chronologie des enquêtes relevant du dispositif « Génération » du CEREQ

	↓	Année d'entrée sur le marché du travail de la cohorte ("Génération")						
		Année d'interrogation						
	Génération 1998	Génération 2001	Génération 2004	Génération 2007	Génération 2010	Génération 2013	Génération 2017	Génération 2021
1998	↓							
1999								
2000								
2001	3 ans	↓						
2002								
2003	5 ans							
2004		3 ans	↓					
2005	7 ans							
2006								
2007			3 ans	↓				
2008	10 ans							
2009			5 ans					
2010				3 ans	↓			
2011			7 ans					
2012								
2013					3 ans	↓		
2014								
2015					5 ans			
2016						3 ans		
2017					7 ans		↓	
2018								
2019								
2020							3 ans	
2021								↓
2022								
2023							6 ans	
2024								3 ans
2025								

1.2-La définition des « primo sortants » du système éducation : un champ évolutif

Les enquêtes « Génération » s'intéressent aux « primo-sortants » du système éducatif,

qui arrêtent leurs études au cours ou à l'issue d'une même année scolaire. Quatre principaux critères cumulatifs déterminent l'appartenance à la « Génération » ayant l'année n comme millésime :

- avoir été inscrit dans un établissement de formation durant l'année scolaire n-1/n ;
- ne pas avoir durablement interrompu ses études avant l'année de sortie, sauf pour raison de santé ou, lorsqu'il existait encore, en raison du service national
- ne pas avoir repris d'études au cours de l'année qui suit l'entrée sur le marché du travail, quel que soit le lieu de l'établissement d'inscription
- avoir au plus 35 ans l'année n.

La façon de définir plus précisément ces critères a évolué dans le temps. Le tableau 1 présente ces évolutions. En particulier, l'allègement des contraintes techniques s'imposant à la collecte ont permis d'élargir le champ géographique des établissements couverts de façon à intégrer les départements d'outre-mer. Les restrictions pesant sur le lieu de résidence de l'enquêté au moment de la collecte ont aussi été progressivement allégées. L'évolution des politiques publiques et des comportements a aussi conduit à des modifications, notamment la prise en compte de la possibilité d'une année de césure à partir de la « Génération 2017 ».

Les jeunes en formation par alternance et le cas des élèves fonctionnaires nécessitent par ailleurs d'arbitrer sur la façon de les considérer. Sont-ils en formation ou en emploi ? Bien que faisant l'objet d'un contrat de travail, les contrats d'apprentissage sont considérés comme relevant de la formation initiale et non comme un emploi. Comme pour les autres formations scolaires ou universitaires, un apprenti qui n'est pas sorti du système éducatif avant l'année scolaire de référence et qui ne poursuit pas d'études ou d'apprentissage ensuite, entre dans le champ de l'enquête. A contrario, les contrats de qualification, lorsqu'ils existaient, étaient considérés comme des emplois parce qu'ils relevaient d'abord de la formation continue et des politiques d'emploi. Jusqu'à la « Génération 2013 », les contrats de professionnalisation ont été traités comme les contrats de qualification. Depuis la « Génération 2017 », ils sont traités comme l'apprentissage. De leur côté, les élèves fonctionnaires, y compris ceux qui le deviennent dans la prolongation de leur formation initiale, étaient considérés comme étant en emploi pendant leur période de formation rémunérée jusqu'à la « Génération 2010 ». Depuis la « Génération 2013 », ils sont considérés comme accédant à l'emploi au moment de leur première prise de poste.

Ces évolutions de champ nécessitent une vigilance particulière lorsque les enquêtes sont exploitées pour étudier des évolutions. Les variables disponibles dans la base

permettent notamment de s'assurer de la cohérence de champ sur les formations de sortie prises en compte, la localisation des établissements de scolarisation l'année de sortie, la nationalité des étudiants.

Tableau 1. Evolution du champ de l'enquête d'une « Génération » à l'autre

	Génération 1998 Génération 2001	Génération 2004 Génération 2007	Génération 2010 Génération 2013	Génération 2017
Localisation de l'établissement d'inscription l'année de sortie	Métropole		France (Métropole & DOM*)	
Période de sortie du système éducatif	Année civile n	D'octobre $n-1$ à octobre n		D'octobre $n-1$ à décembre n
Durée possible d'interruption des études avant l'année de sortie	< 12 mois			< 16 mois
Durée minimale sans reprise d'étude après l'année de sortie	12 mois			16 mois
Prise en compte des sortants de 6ème et 5ème	Oui	Non		
Prise en compte des étudiants étrangers sortants de l'Université	Non	Oui		
Prise en compte de certaines sorties de contrats de professionnalisation	Non			Oui
Lieu de résidence de l'enquêté à la date d'enquête	Métropole		France	Aucune restriction

* Guadeloupe, Guyane, La Réunion, Martinique, Mayotte.

2. La structure du questionnaire : des modules standards et des variations thématiques

Les questionnaires comprennent des modules standards, disponibles pour toutes les cohortes, dont la formulation et le degré de détail peuvent différer d'une « Génération » à l'autre. Des modules thématiques, variables d'une enquête à l'autre, complètent les modules standards. Certains résultent d'initiatives propres au CEREQ, liées à son programme de recherche. D'autres constituent des « extensions de questionnement » destinées à répondre à des besoins de partenaires extérieurs. Ces modules thématiques ne sont pas systématiquement renvoyés à la fin du questionnaire standard. Ils s'articulent avec lui de façon à assurer la meilleure fluidité possible lors de la passation du questionnaire.

Dans cette partie, nous présentons principalement les modules standards. Certains ne sont posés que lors de l'interrogation à trois ans, comme la vérification de l'appartenance au champ de l'enquête, la description du parcours scolaire et la caractérisation socio-démographique des enquêtés. Le calendrier d'activité et les modules qui s'adosent à lui pour décrire les emplois successifs occupés et les périodes de non emploi constituent le cœur de l'enquête de l'interrogation à trois ans, comme des suivantes.

2.1-La vérification de l'appartenance au champ de l'enquête

Lors de l'interrogation à trois ans, le questionnaire commence systématiquement par un module destiné à vérifier l'appartenance de l'enquêté au champ de l'enquête :

validation de l'identité du répondant, de son niveau de sortie et vérification des critères d'appartenance au champ, en particulier l'absence d'interruption des études avant l'année scolaire de référence et l'absence de reprise d'étude ensuite. A titre d'exemple, pour la « Génération 2004 », 44% des personnes ayant accepté de répondre à l'enquête se sont avérées hors champs lors de cette étape (Aliaga, Duploux et al., 2010).

2.2-La description du parcours scolaire

Le questionnaire revient ensuite sur le parcours scolaire de la personne enquêtée : redoublement avant la sixième, nature de la classe de troisième, classe suivie après la troisième, conformité de cette orientation avec ses choix, nature du baccalauréat obtenu s'il y a lieu, mention au baccalauréat, classe suivie après le baccalauréat, conformité de cette orientation avec ses choix. Le diplôme de sortie et, quand il est différent, le plus haut diplôme atteint, sont définis très précisément. L'enquête recense les autres diplômes réussis dans l'enseignement supérieur sans en détailler la spécialité.

Le module consacré au parcours scolaire aborde aussi les expériences professionnelles en cours d'études et les motifs d'arrêt des études.

2.3-Le calendrier d'activité et les modules associés

Le « calendrier d'activité » constitue le cœur du questionnaire, à la première interrogation et aux suivantes pour les cohortes interrogées plusieurs fois. Il décrit l'évolution de la situation d'activité de l'enquêté de sa sortie de formation initiale à la date d'enquête : les périodes d'emploi, les périodes de recherche d'emploi, les périodes de reprise d'études, les périodes de formation et les périodes dans une autre situation. Le schéma 2 présente un exemple de calendrier renseigné.

Schéma 2. Exemple de remplissage d'un calendrier

	année n												année n+1												année n+2												année n+3																												
	jan	fev	mars	avril	mai	juin	juil	août	sept	oct	nov	déc	jan	fev	mars	avril	mai	juin	juil	août	sept	oct	nov	déc	jan	fev	mars	avril	mai	juin	juil	août	sept	oct	nov	déc	jan	fev	mars	avril	mai	juin	juil																						
Interim																																																																	
Entreprise													EMPLOYEUR 1						EMPLOYEUR 2																																														
Recherche d'emploi																																																																	
Reprise d'études																																																																	
Formation																																																																	
Autres situations																																																																	

Le mois calendaire sert d'unité de temps : une seule situation d'activité peut être enregistrée pour un mois donné. Pour les emplois, tout changement d'employeur donne lieu à l'enregistrement d'une nouvelle période d'emploi. Un changement d'établissement au sein d'une même entreprise est traité comme un changement d'employeur.

L'intérim est traité de façon spécifique et une question permet d'isoler les « jobs d'été » effectués l'année de sortie du système éducatif (ils ne sont pas caractérisés ensuite à la différence des autres périodes d'emploi). La situation de recherche d'emploi est

déclarative. Elle est indépendante du fait d'être inscrit comme demandeur d'emploi à l'Agence nationale pour l'emploi, devenue Pole Emploi fin 2008. Les reprises d'études supposent une inscription à temps plein dans un établissement scolaire ou universitaire.

Le calendrier est renseigné de façon chronologique en privilégiant les situations d'emploi puisque, pour toute nouvelle période à décrire, l'enquêteur commence par demander si la personne enquêtée a travaillé au cours du mois.

Une fois que le calendrier est totalement rempli, un traitement automatique qualifie chaque période décrite en fonction de trois critères : la situation d'activité, la durée (période « courte » ou « longue ») et la contemporanéité (période encore active à la date d'enquête ou période du passé). En fonction de ces qualifications, des modules de caractérisation des périodes s'ouvrent ou non, avec plus ou moins de détail, selon des configurations qui peuvent changer d'une enquête à l'autre. Les situations actives à la date d'enquête donnent toujours lieu à la description la plus complète. Les périodes « courtes » passées de recherche d'emploi, de formation ou dans une « autre » situation ne sont pas décrites, pas plus que les jobs d'été.

Les périodes d'emploi sont généralement décrites de façon détaillée. Le nom de l'employeur est collecté pour disposer du secteur d'activité et de la taille d'entreprises. Plusieurs questions sont posées pour pouvoir coder la profession au niveau le plus fin de la nomenclature des catégories socioprofessionnelles de l'INSEE. L'enquête questionne aussi sur le contrat de travail, le temps de travail et la rémunération. Afin de permettre d'étudier assez finement la mobilité professionnelle en début de carrière, le questionnaire enregistre la profession, le type de contrat de travail et la rémunération à l'embauche et en fin de période d'emploi pour les périodes d'emploi du passé. Les raisons d'interruption de l'emploi sont également questionnées.

2.4-Les caractéristiques socio-démographiques

Quelques caractéristiques socio-démographiques sont collectées lors de l'interrogation à trois ans afin de permettre l'étude de leurs effets éventuels sur les conditions d'entrée sur le marché du travail. Elles permettent aussi des analyses sur les inégalités sociales et sur les discriminations potentielles.

Les enquêtes disposent ainsi du sexe, du pays de naissance et de la nationalité de la personne enquêtée. Elles collectent le pays de naissance et la nationalité de ses parents, leur situation d'activité à la fin des études de l'enquêté et leurs professions s'ils ont déjà travaillé. Celle-ci n'est renseignée qu'à grosse maille, selon la nomenclature de l'INSEE codée sur une seule position (ouvrier, employé, professions intermédiaires, etc.). La nationalité de l'enquêté est collectée de façon binaire (française ou étrangère). Pour les parents, l'enquête est plus précise en distinguant aussi les Français par acquisition et les doubles nationalités. A partir de la « Génération 2010 », le questionnaire aborde les langues parlées par les parents durant l'enfance.

Pour certaines cohortes, le questionnaire a également abordé l'existence de frères,

sœurs ou d'un conjoint, ainsi que le niveau de diplôme de parents. A contrario, le questionnaire de la « Génération 2007 » a été très allégé. Certaines informations habituellement collectées ne l'ont pas été.

Pour suivre les mobilités ou étudier l'existence d'effets de territoire, des informations sont disponibles sur le lieu de résidence en sixième, au baccalauréat, à la fin des études et à la date d'enquête. Pour la fin des études et la date d'enquête, la disponibilité de l'adresse permet, après géocodage, d'identifier l'habitat dans des quartiers relevant des politiques de la ville. En dehors de ces cas, l'information disponible identifie la région, le département, voire la zone d'emploi, ainsi que le type de commune selon la typologie territoriale de l'Insee en vigueur au moment de la production des fichiers (zonage en aires urbaines, zonage des aires d'attraction des villes). L'identification des communes est couverte par le secret statistique.

2.5. Les questions subjectives sur les perspectives professionnelles et la satisfaction dans l'emploi

Dès sa première recommandation, le rapport Stiglitz-Sen-Fitoussi sur la mesure des performances économiques et du progrès sociale (Stiglitz, Sen et al., 2009) invite les services statistiques à « intégrer dans leurs enquêtes des questions visant à connaître l'évaluation que chacun fait de sa vie, ses expériences gratifiantes et ses priorités ». Dès la « Génération 1998 », le questionnaire intègre cette préoccupation en abordant les perspectives professionnelles et la satisfaction dans l'emploi. Elles permettent, par exemple, de proposer un angle d'approche complémentaire aux analyses qui s'intéressent au déclassement à partir de grilles « objectives » construites en utilisant les nomenclatures de diplômes et de professions. Certaines de ces questions sont posées à tous. D'autres ne le sont qu'aux personnes en emploi à la date d'enquête.

A l'exception de l'enquête « Génération 2007 », les enquêtes abordent aussi le sentiment de discrimination. Jusqu'à la « Génération 2013 », il est abordé à partir de trois questions : le sentiment d'avoir été discriminé au moins une fois « à l'embauche », la fréquence (une fois, plusieurs fois, souvent) et les motifs (nom, genre, couleur de peau, accent, lieu de résidence, look, ...). Pour la « Génération 2017 », la formulation de la question initiale élargit le champ en évoquant la discrimination « à l'embauche ou dans l'emploi » mais les questions sur la fréquence et les motifs ne sont pas posées.

2.6-Les modules thématiques

Lors de certaines enquêtes, des modules spécifiques sont développés pour éclairer des points particuliers de la relation entre la formation initiale et les conditions d'accès à l'emploi, en lien avec des projets de recherche du CEREQ ou, le plus souvent, pour répondre à des besoins spécifiques de partenaires institutionnels, tels que des services statistiques ministériels, Pole Emploi ou l'Association de gestion du fonds pour l'insertion professionnelle des personnes handicapées. Ces modules peuvent se limiter à quelques questions mais sont parfois plus étoffés. Cette partie en présente quelques-uns succinctement de manière illustrative. Les documents disponibles avec les fichiers

de données fournissent des informations plus complètes.

Ainsi, l'enquête « Génération 2004 » à trois ans comprend un module spécifique aux apprentis qui aborde les relations avec l'entreprise d'accueil (conditions de travail, rôle dans la formation, recrutement à l'issue de la formation). Un module destiné aux thésards aborde leur projet professionnel initial, leur mode de rémunération, leur situation postdoctorale. Un module repère les jeunes ayant un problème de santé ou de handicap et la reconnaissance de ce handicap.

L'enquête « Génération 2007 » à trois ans intègre à nouveau le module destiné aux thésards. Par ailleurs, un module assez complet porte sur les stages faits en cours de scolarité. Il recense d'abord l'ensemble des stages d'un mois ou plus réalisés en cours de scolarité puis en échantillonne certains pour les décrire précisément^[3] (existence d'une convention, rémunération, lien avec la formation, intensité du suivi par le maître de stage...). Ce module n'est cependant passé qu'auprès d'un sous-échantillon de l'échantillon principal. Il donne lieu à une table spécifique. Dans la même enquête, un module s'intéresse au recours aux intermédiaires du marché du travail, lui aussi destiné à un sous-échantillon de l'enquête.

L'enquête « Génération 2010 » à trois ans comprend des modules qui abordent les séjours à l'étranger, la professionnalisation dans l'enseignement supérieur, l'existence d'une situation de handicap, ainsi que le module destiné aux thésards.

L'enquête « Génération 2013 » à trois ans comprend à nouveaux les modules consacrés aux situations de handicap et aux thésards. Des questions développent aussi les séjours à l'étranger, le rapport au travail des jeunes, l'attractivité de la fonction publique, l'entrepreneuriat et le passage par certains dispositifs d'accompagnement (garantie jeune, service civique, contrats d'insertion dans la vie sociale).

Dans l'enquête « Génération 2017 », des modules portent sur les conditions de logement au cours de la dernière année de formation, les séjours à l'étranger, le service civique, les risques physiques et chimiques et l'attractivité de la fonction publique.

3. De la base de sondage à la base de résultats

Les enquêtes « Génération » sont produites par le département des entrées et évolutions dans la vie active du CEREQ, composé d'une quinzaine à une vingtaine de personnes selon les périodes, pour les enquêtes présentées. En son sein, une équipe de gestion d'enquêtes, d'une demi-douzaine de personnes, est spécifiquement chargée de la production des enquêtes, de la constitution de la base de sondage à la production des fichiers finaux. Elle est principalement composée d'ingénieurs et personnels techniques, de recherche et de formation et placée sous la responsabilité d'un statisticien de l'INSEE. Elle comprend également une personne méthodologue.

Les autres membres du département sont économistes, sociologues ou statisticien de formation et, statutairement, ingénieurs d'études, ingénieurs de recherches ou

statisticiens de l'Insee. Ils participent, par leurs travaux, aux exploitations de l'enquête, mais aussi à l'élaboration du questionnaire. Celui-ci fait également l'objet d'échanges au sein d'un comité de concertation composé à cet effet, avec des représentants d'acteurs sociaux, notamment des syndicats salariés, et des partenaires du CEREQ, dont le service statistique ministériel de l'éducation nationale, celui de l'enseignement supérieur et celui du travail.

Pour certaines enquêtes, l'équipe de gestion des enquêtes a publié des documents méthodologiques qui présentent les différentes étapes de production, de la base de sondage à la production des fichiers de résultats. C'est notamment le cas pour la première interrogation de la « Génération 2004 » (Aliaga, Duploux, et al., 2010), la première interrogation de la « Génération 2007 » (Rouaud, 2012), la première interrogation de la « Génération 2010 » (Barret, Dzikowski et al., 2019), sa deuxième interrogation (Olaria, Barret et al., 2019) et sa troisième interrogation (Cisse, Gaubert et al., 2019).

Les dictionnaires des codes, associés aux fichiers de diffusion, résument également le processus de production des enquêtes abordé dans cette partie.

3.1. Contourner l'absence de base de sondage correspondant au champ de l'enquête

Il n'existe pas de fichiers centralisés de l'ensemble des élèves et étudiants inscrits dans les établissements d'enseignement publics et privés français une année donnée et a fortiori, de fichiers de sortants de formation initiale. La réalisation des enquêtes « Génération » suppose donc de construire en amont une base de sondage spécifique en collectant les listes d'élèves et d'étudiants, inscrits ou sortants des établissements de l'enseignement secondaire et de l'enseignement supérieur au cours de l'année scolaire de référence.

Pour certaines composantes, il est possible de disposer de fichiers centralisés. Jusqu'à la « Génération 2013 », les rectorats transmettaient ainsi au CEREQ les informations concernant les collèges, les lycées et BTS relevant du ministère de l'éducation nationale dans leur académie, y compris les établissements privés sous contrat. A partir de la « Génération 2017 » une refonte de système d'information « Scolarité » permet au service statistique ministériel de l'éducation nationale de fournir au CEREQ un fichier national unique pour ce champ. Le développement de systèmes d'informations centralisés permet aussi de collecter centralement les inscriptions d'autres composantes, comme les centres de formation d'apprentis ou les lycées agricoles à partir de la « Génération 2013 ». Pour les composantes non couvertes centralement, notamment les universités, les écoles de commerce et les écoles d'ingénieurs, le CEREQ collecte directement auprès des établissements concernés des extraits de leurs fichiers de gestion. La liste des établissements à interroger est définie à partir d'un répertoire national géré par le ministère de l'éducation nationale.

Ces informations sont collectées pour l'année scolaire de référence et pour la suivante,

de façon à pouvoir identifier, quand c'est possible, les cas de poursuite d'études, en s'appuyant sur les informations d'état-civil disponibles, notamment les noms, prénoms et année de naissance, ou sur l'identifiant national des élèves et des étudiants, l'INE – cet identifiant de gestion est supposé suivre l'élève tout au cours de sa scolarité, y compris dans l'enseignement supérieur, mais le CEREQ ne le récupère que pour une partie des établissements couverts par ses enquêtes.

A l'issue de la collecte des fichiers d'inscrits, la base de sondage est apurée en supprimant les poursuites d'études repérées et les doublons éventuels en cas d'inscriptions multiples. Une partie des situations de poursuite d'études n'est pas repérable et il n'est pas possible de s'assurer dès cette phase du respect d'autres critères, notamment l'absence d'interruption des études avant l'année scolaire de référence. L'appartenance effective au champ des enquêtés ne peut donc être vérifiée que lors de la passation du questionnaire. Le calibrage de l'échantillon tient compte de l'existence d'un taux de non-appartenance au champ de l'enquête.

3.2. La constitution de l'échantillon

L'échantillon est tiré aléatoirement, avec des probabilités inégales, dans la base de sondage. Le tirage est stratifié par grands niveaux de formation (formation non diplômante, CAP, baccalauréat général, etc.). Au sein de chaque strate, le tirage est équilibré au sens de Deville et Tillé (2004) en tenant compte de la décomposition des effectifs selon différents critères, notamment le sexe et le type de formation (par exemple, pour distinguer les spécialités industrielles et les spécialités tertiaires), mais aussi dans certains cas, d'autres informations comme le fait de résider dans un quartier de politique de la ville ou la région.

Pour la partie de l'échantillon destinée à répondre aux besoins du CEREQ, la stratification vise à disposer d'un nombre de répondants suffisants pour proposer des indicateurs nationaux par grands niveaux de formation, selon la nomenclature ad hoc construite par le CEREQ pour ses publications. Celle-ci s'est d'abord appuyé sur les niveaux de sortie, avant de privilégier le plus haut diplôme atteint (13 niveaux pour la « Génération 2017 »). Pour les cohortes qui font l'objet de plusieurs interrogations, l'échantillon initial est plus important que pour les cohortes interrogées une seule fois afin de tenir compte de l'attrition entre deux interrogations. Pour ces grosses cohortes, l'échantillon est aussi construit pour permettre de proposer des informations par région avec une nomenclature de niveau plus agrégée que pour les résultats nationaux.

Comme pour toute enquête par sondage, la précision des résultats dépend de la taille de l'échantillon mais aussi du plan de sondage. Ce point est particulièrement important dans le cas des enquêtes « Génération » dans la mesure où la taille de l'échantillon peut être fortement grossie pour répondre à des demandes de partenaires qui ne s'intéressent qu'à des sous-populations particulières donnant lieu à des extensions d'échantillons ciblées. L'annexe liste ces extensions. Elle présente également le nombre de répondants visés.

Le calibrage de l'échantillon est effectué en partant du nombre de répondants cibles et d'hypothèses sur les taux de réponse et les taux de hors champ. Ces taux sont différents selon les niveaux de formation et types d'établissements. Ils sont estimés à partir des observations faites lors des enquêtes précédentes.

Le calibrage de l'échantillon suppose aussi de faire des hypothèses sur la capacité à retrouver les personnes trois ans après leur sortie de formation initiale, à partir des coordonnées disponibles dans les fichiers de gestion du dernier établissement qu'ils ont fréquenté - ces coordonnées (adresse, téléphone) peuvent être les leurs, celles de leurs parents ou concerner leur logement étudiant. De plus, ces coordonnées ne sont pas toujours disponibles. Une fois l'échantillon tiré, il est enrichi par une recherche de coordonnées téléphoniques sur la base des noms, prénoms et adresses. Cette phase était d'autant plus efficace que le patronyme est peu répandu ou la commune de résidence de petite taille. Avec le temps, l'efficacité de la phase d'enrichissement s'est réduite mais le développement des téléphones portables permet de disposer de davantage de coordonnées à jour. A titre d'exemple, une adresse complète était disponible pour 88% des individus de la base de sondage de l'enquête « Génération 2017 » ; un numéro de téléphone d'élève ou de parent pour 80% ; un mél d'élève pour 49%[\[4\]](#).

L'échantillon est enfin divisé en deux parties, par tirage aléatoire, pour distinguer un échantillon principal et un échantillon de réserve. Si les effectifs cibles sont atteints ou dépassés avec l'échantillon principal, l'échantillon de réserve n'est pas mobilisé, ce qui permet de contenir les coûts de collecte. En cas de non atteinte des objectifs, l'engagement de la réserve s'effectue par strate. A titre d'exemple, pour la « Génération 2004 », 403 476 individus ont été échantillonnés pour une cible de 65 000 répondants, extensions d'échantillon incluses. 90% de cet échantillon alimentait l'échantillon principal (Aliaga, Duploux, et al., 2010).

3.3. La collecte

La collecte est réalisée par un prestataire de service spécialisé. Jusqu'à la « Génération 2013 », c'est une collecte par téléphone assistée par ordinateur. A partir de la « Génération 2017 », la collecte devient bimodale, par Internet et par téléphone. Cette évolution permet d'élargir le champ de l'enquête aux jeunes à l'étranger à la date d'enquête. Elle induit aussi des effets de mode sur certaines questions, notamment des questions d'opinion (Dabet, Mazari, 2023). Dans les fichiers de résultats, une variable renseigne sur le mode de collecte. Il est donc possible d'examiner ce risque.

L'annexe présente le nombre de répondants par enquête, en distinguant la cible relevant de l'échantillon du CEREQ de celles relevant des extensions d'échantillon.

3.4. Les traitements aval

Une fois la collecte réalisée, les fichiers sont apurés en appliquant des contrôles de cohérence, des redressements et des recodages de certaines variables.

En particulier, lors de cette phase, la profession exercée au cours des différentes périodes d'emploi recensées est codée selon la nomenclature la plus fine des professions et catégories socioprofessionnelles, en s'appuyant sur les algorithmes de l'INSEE et en se référant à la nomenclature disponible la plus récente au moment de la production des fichiers. Le secteur d'activité est codé dans la nomenclature d'activité de l'INSEE en partant du nom de l'employeur collecté et de son secteur d'activité déclaré dans le répertoire Sirene, géré par l'INSEE. Les informations géographiques sont codées en s'appuyant sur le code officiel géographique et les zonages d'études de l'INSEE (zones d'emploi, typologie associée aux zonages en aires urbaines et en aires d'attraction des villes). Un géocodage des adresses permet de repérer la résidence dans un quartier de politique de la ville. Les niveaux de sortie et les plus hauts diplômes atteints sont agrégés selon une nomenclature ad hoc du CEREQ évolutive, en s'appuyant sur les niveaux de formations et spécialités détaillées, collectés dans la base de sondage et lors de la passation du questionnaire. Enfin, les variables de rémunération font l'objet de contrôles de cohérence internes pour identifier et corriger les points aberrants.

La pondération finale est calculée en plusieurs étapes pour tenir compte des poids de tirage par strate, de la capacité à disposer des coordonnées, de l'acceptation à répondre, de la probabilité d'être hors champs. Un calage sur marges est aussi réalisé en prenant comme référence les effectifs de sortants par grands niveaux estimés à partir de l'enquête emploi de l'INSEE. Aliaga, Duploux et al. (2010) et Rouaud (2012) présentent de façon détaillée le calcul.

3.5. Les fichiers de résultats

Chaque enquête donne lieu à la mise à disposition de trois fichiers : une table « individus », une table des séquences d'emploi et une table des séquences de non emploi. Pour la « Génération 2017 », ces deux dernières tables sont fondues en une seule table de séquences.

La table « individus » comprend une ligne par répondant et toutes les informations qui ne sont pas liées à une séquence particulière du calendrier d'activité : les caractéristiques sociodémographiques, le parcours scolaires, les expériences professionnelles pendant la scolarité, les perspectives professionnelles, etc. La table comprend aussi des variables synthétiques construites à partir du calendrier, comme la situation d'activité mois par mois ou le nombre de mois en emploi.

La table des séquences d'emploi comprend une ligne par période d'emploi décrite dans le calendrier, avec toutes les caractéristiques de cet emploi. Les réponses aux questions d'opinion sur l'adéquation de l'emploi occupé avec sa formation ou sa juste rémunération sont dans cette table, associées à l'emploi actif à la date d'enquête.

La table des séquences de non emploi est construite de manière analogue.

Un identifiant non significatif associé au répondant permet de rapprocher les séquences

d'emploi et de non emploi avec les informations de la table « individus ».

Pour les cohortes faisant l'objet de plusieurs interrogations, les fichiers de résultats des ré-interrogations sont limités au champ des enquêtés ayant répondu à la ré-interrogation concernée mais ils intègrent les réponses des interrogations précédentes.

Les répondants primo sortants relevant des extensions d'échantillon sont généralement intégrés dans les fichiers de résultats mis à disposition. Les autres répondants, parce qu'ils n'entrent pas dans le champ de l'enquête CEREQ, ne sont pas intégrés dans ces tables.

4. Conditions d'accès aux données

4.1 Le label « Statistique publique » et ses conséquences pour l'accès aux données

Les enquêtes « Génération » sont labellisées « statistique publique » au sens de la loi n°51-711 du 7 juin 1951 sur l'obligation, la coordination et le secret en matière de statistiques. A ce titre, leur mise en œuvre nécessite de respecter des procédures mises en place pour garantir l'intérêt, la pertinence et la qualité des enquêtes. Leur opportunité est discutée au sein du Conseil national de l'information statistique et leur méthodologie est examinée par le Comité du Label. Cet examen porte sur tous les aspects : base de sondage, méthode d'échantillonnage, redressement de la non réponse totale, formulation et enchaînement des questions, méthode de passation du questionnaire.

L'obtention du label « statistique publique » permet de réduire le taux de non réponse. Il induit aussi une obligation d'accessibilité des données aux chercheurs sous réserve de respecter certaines règles destinées à garantir le respect du secret statistique.

4.2 Modalités d'accès aux données

Les fichiers de résultats sont systématiquement mis à disposition des chercheurs sur la plateforme PROGEDO après une période d'expertise et de consolidation des données. Ils sont accessibles aux chercheurs sur la base d'un projet de recherche défini et après signature d'un engagement à respecter certaines règles de sécurité et de confidentialité.

Le tableau 2 liste les enquêtes « Génération » disponibles sur la plateforme en précisant leur année de collecte et leur DOI. Le tableau figurant en annexe présente pour chacune d'elles le nombre de répondants et les objectifs ayant guidé le tirage de l'échantillon.

Tableau 2. DOI des enquêtes « Génération » disponibles sur PROGEDO

<https://data.progedo.fr/series/adisp/generation>.

Année de collecte	Nom de l'enquête	DOI
2001	Génération 1998 - Interrogation à 3 ans	doi:10.13144/lil-0167
2003	Génération 1998 - Interrogation à 5 ans	doi:10.13144/lil-0245
2004	Génération 2001 - Interrogation à 3 ans	doi:10.13144/lil-0248
2005	Génération 1998 - Interrogation à 7 ans	doi:10.13144/lil-0370
2007	Génération 2004 - Interrogation à 3 ans	doi:10.13144/lil-0474
2008	Génération 1998 - Interrogation à 10 ans	doi:10.13144/lil-0543
2009	Génération 2004 - Interrogation à 5 ans	doi:10.13144/lil-0575
2010	Génération 2007 - Interrogation à 3 ans	doi:10.13144/lil-0629
2011	Génération 2004 - Interrogation à 7 ans	doi:10.13144/lil-0952
2013	Génération 2010 - Interrogation à 3 ans	doi:10.13144/lil-1084
2015	Génération 2010 - Interrogation à 5 ans	doi:10.13144/lil-1292
2016	Génération 2013 - Interrogation à 3 ans	doi:10.13144/lil-1293
2017	Génération 2010 - Interrogation à 7 ans	doi:10.13144/lil-1439

L'interrogation à trois ans de la « Génération 2017 » doit être mise à disposition sur la plateforme début 2024. L'interrogation à six ans de cette « Génération » a été collecté au dernier trimestre 2023. Elle devrait être disponible fin 2024 ou début 2025. L'interrogation à trois ans de la « Génération 2021 » est actuellement en préparation pour une collecte programmée à l'automne 2024.

Dans les fichiers mis à disposition par PROGEDO, le CEREQ agrège certaines informations qu'il collecte de façon détaillée pour éviter que le croisement d'informations ne rompe le secret statistique. C'est notamment le cas des informations géographiques : la commune n'est pas fournie pour les lieux de résidence ; les pays de naissance des parents sont regroupés par grandes zones géographiques. L'identification des employeurs n'est pas non plus livrée dans les fichiers proposés par PROGEDO.

Conclusion

Les enquêtes « Génération » du CEREQ sont un outil adapté pour analyser statistiquement les premiers pas dans la vie active des jeunes à la sortie de formation initiale, avec une approche multidimensionnelle de l'insertion et dans une optique comparative entre niveaux, spécialités et modes de formation.

Elles permettent aussi d'aborder, par des analyses contrôlées, les inégalités et discriminations sociales selon le genre, l'origine sociale, l'ascendance migratoire ou le lieu de résidence. Elles peuvent aussi permettre d'aborder les métiers, les secteurs, les taille d'entreprises, qui recrutent des jeunes et leur capacité à les retenir.

La taille de l'échantillon limite cependant la finesse des analyses. En particulier, ces enquêtes ne sont pas construites pour évaluer les performances de formations ou d'un territoire précis. L'utilisation de ces enquêtes nécessitent aussi d'être attentifs à deux

points : premièrement, l'évolution du champ de l'enquête oblige à une certaine vigilance si l'on souhaite réaliser des comparaisons intertemporelles. Deuxièmement, la taille de l'échantillon ne suffit pas à déduire le degré de finesse des analyses possibles car une partie résulte d'un sur-échantillonnage ciblé. Il est donc nécessaire de regarder comment l'échantillon a été construit.

Les enquêtes ont donné lieu à des dizaines de publications[5], principalement dans les collections du CEREQ. On peut s'y reporter pour avoir des exemples d'utilisation.

En particulier, chaque enquête donne lieu à une publication rapide de premiers résultats dans la collection des Bref du CEREQ, la plus récente étant Couppié, Gaubert et al. (2022) pour la « Génération 2017 ». Pour les enquêtes à trois ans, cette première publication est suivie de résultats plus complets dans un ouvrage dédié intitulé Quand l'école est finie. Ces ouvrages proposent des fiches thématiques de tableaux et graphiques accompagnés de commentaires synthétiques qui balayent l'ensemble du questionnaire, comme Lopez, Epiphane et al. (2002) pour la « Génération 1998 » ou Le Bayon, Dabet et al. (2022) pour la « Génération 2017. Les sept premières années de vie actives ont aussi donné lieu à des publications étoffées mais moins standardisées pour la « Génération 1998 » (Couppié, Gasquet et al., 2006), la « Génération 2004 » (Coupprie, H., Dzikowzski, 2015) et la « Génération 2010 » (Couppié, Dupray et al. 2022). Des groupes d'exploitation sont aussi mis en place par le CEREQ avec des appels à proposition ouverts aux chercheurs externes. Les études menées dans ce cadre donnent lieu à des publications groupées, comme (Calmand, Couppié et al., 2017) pour la « Génération 2010 ». Enfin, les enquêtes déposées sur la plateforme PROGEDO ont fait l'objet, en moyenne, de 70 demandes d'accès chacune, avec une étendue allant de 41 à 97 demandes d'accès, selon les éditions (chiffres arrrêtés à fin 2023).

Références bibliographiques

Aliaga C., Duplouy, B., Jugnot, S., Ryk, F. (2010), « Enquête Génération 2004, Méthodologie et bilan de la 1ère interrogation ». Net.doc, n°63. CEREQ.

Barret, C., Dzikowski, C., Mazari, Z., Rouaud, P., Ryk, F., Volle, N. (2019). « Enquête Génération 2010. 1ère interrogation, printemps 2013. Bilan méthodologique ». Cereq Etudes, n°25. CEREQ.

Calmand, J., Couppié, T., Henrard, V. (2017). « Rendement éducatif, parcours et inégalités dans l'insertion des jeunes. Recueil d'études sur la Génération 2010 ». CEREQ Echanges, n° 5. CEREQ

Cisse, M., Gaubert, E., Mazari, Z., Olaria, M., coord. (2019). « Enquête Génération 2010. 3e interrogation, automne 2017. Bilan méthodologique ». Cereq Etudes, n°27. CEREQ.

Couppié, T., Dieusaert, P., Vignale, M. (2020). « Que deviennent les jeunes des quartiers prioritaires de la ville après leur bac ? ». Bref, n°391. CEREQ.

Couppié, T., Dupray, A., Gasquet, C., Personnaz, E., coord. (2022). « Chemins vers l'emploi et la vie adulte : l'inégalité des Possibles. Les sept premières années de vie active de la Génération 2010 ». Cereq Essentiels, n°4. CEREQ.

Couppié, T., Epiphane, D. (2019). « Et les femmes devinrent plus diplômées que les hommes... ». Bref, n°373. CEREQ.

Couppié, T., Gaubert, G., Personnaz, E. (2022). « Enquête 2020 auprès de la Génération 2017. Des parcours contrastés, une insertion plus favorable, jusqu'à... ». Bref, n°422. CEREQ.

Couppié, T., coord, Gasquet, C., coord., Lopez, A., coord. (2006). Quand la carrière commence, Les sept premières années de vie active de la Génération 98. CEREQ.

Coupprie, H., coord., Dzikowzski, C., coord. (2015), « Les sept premières années de vie active de la Génération 2004. Enquêtes 2007, 2009 et 2011 ». Net.doc, n°139. CEREQ.

Dabet, G., Mazari, Z., Oujna, I. (2023). « L'effet de mode dans les enquêtes multimodes (internet/téléphone). Estimation et décomposition ». Cereq Etudes, n°45. CEREQ.

Deville, J.-C., Tillé, Y. (2004). « Efficient Balanced Sampling : The Cube Method ». Biometrika, Vol 91, No 4, pp 893-912.

Jugnot, S. (2023). « L'accès à l'emploi des immigrés et enfants d'immigrés de la Génération 2017 », Bref, n°434. CEREQ.

Le Bayon, F., coord., Dabet G., coord., Joseph, O., coord., Olaria, M., coord. (2022). Quand l'Ecole est finie. Premiers pas dans la vie active de la Génération 2017. CEREQ.

Lopez, A., coord., Epiphane, D., coord. (2002). Quand l'Ecole est finie. Premiers pas dans la vie active de la Génération 1998. CEREQ.

Rouaud, P. (2012) « Enquête 2010 auprès de la Génération 2007. Bilan des redressements ». Net.doc, n°95. CEREQ.

Stiglitz, E., Sen, A., Fitoussi, J.-P. (2009). Richesse des nations et bien-être des individus. Rapport de la Commission sur la mesure des performances économiques et du progrès social. Editions Odile Jacob.

Olaria, M., Barret, C., Cisse, M., Dzikowski, C., Gaubert, E., Mazari, Z., Ryk, F. (2019). « Enquête Génération 2010. 2ème interrogation, printemps 2015. Bilan méthodologique ». Cereq Etudes, n°26. CEREQ.

Vallet, L.-A., Goffette, C. (2022). « De la génération 1992 à la génération 2010, comment le lien entre la qualité de l'insertion professionnelle des jeunes et leur origine socio-culturelle a-t-il évolué ? ». <https://sciencespo.hal.science/hal-03618457/>

Annexe - Nombre de répondants et cibles de l'échantillonnage

Nom de l'enquête (en grisé, les cohortes interrogées plusieurs fois)	Nombre de répondants dans le champ "CEREQ" (voir partie 1.2)	Objectifs visés lors de la construction de l'échantillon (les exploitations nationales visent des résultats nationaux par grands niveaux de formation, voir partie 3.2)
Génération 1998 à 3 ans	55 300	Disposer d'environ 30 000 répondants pour les exploitations nationales du CEREQ, 10 000 répondants supplémentaires pour permettre une exploitation régionale par grands niveaux de formation et 15 000 répondants pour les extensions partenariales ciblées (sortants des formations de la santé, de doctorat, sortants des établissements des régions Aquitaine, Île-de-France, Lorraine, Provence-Alpes-Côte d'Azur et Rhône-Alpes).
Génération 1998 à 5 ans	22 000	La 2e interrogation a concerné les répondants de la 1ère interrogation destinés aux exploitations nationales du CEREQ et à l'extension relative aux sortants des formations sanitaires.
Génération 1998 à 7 ans	16 000	La 3e interrogation a concerné les répondants de la 2e interrogation.
Génération 1998 à 10 ans	11 200	La 4e interrogation a concerné les répondants de la 3e interrogation.
Génération 2001 à 3 ans	25 000	Disposer d'environ 10 000 répondants pour les exploitations nationales du CEREQ et 15 000 répondants pour les extensions partenariales ciblées (sortants de formations de santé, de doctorat, de licences professionnelles, des filières STAPS, des formations professionnelles du BTP, sortants des établissements des régions Bretagne, Franche-Comté et Lorraine).
Génération 2004 à 3 ans	65 000	Disposer d'environ 35 000 répondants pour les exploitations nationales du CEREQ et 30 000 répondants pour les extensions partenariales ciblées, dont certaines couvrent aussi des sortants "post-initiaux", qui avaient déjà interrompu au moins un an leurs études (Sortants de Licences professionnelles, de L2, de M3, de doctorat, de formations sanitaires et sociales, des formations de la Culture, des formations Jeunesse et Sports, des formations de l'environnement, des formations professionnelles de l'automobile, du BTP et du Commerce, sortants des établissements des régions Basse-normandie, Lorraine, Nord-Pas-de-Calais, Poitou-Charente, Provence-Alpes-Côte d'Azur).
Génération 2004 à 5 ans	20 100	La 2e interrogation a concerné les répondants à la 1ère interrogation destinés aux exploitations nationales du CEREQ et aux extensions relatives aux sortants des formations sanitaires et des formations professionnelles de l'automobile.
Génération 2004 à 7 ans	12 400	La 3e interrogation a concerné les répondants de la 2e interrogation destinés aux exploitations nationales du CEREQ et aux extensions relatives aux sortants des formations sanitaires.
Génération 2007 à 3 ans	24 900	Disposer d'environ 3 000 répondants pour les exploitations nationales du CEREQ et 27 000 répondants pour les extensions partenariales ciblées, dont certaines couvrent aussi des sortants "post-initiaux", qui avaient déjà interrompu au moins un an leurs études (sortants non diplômés, sortants des expérimentations du baccalauréat professionnel à trois ans, de trois filières professionnelles, de l'enseignement supérieur, des formations sanitaires et sociales, des formations Jeunesse et Sports, des formations de l'environnement, des écoles de la fonction publique, sortants des zones urbaines sensibles, sortants des établissements des régions Aquitaine, Bretagne, Franche-Comté, Nord-Pas-de-Calais, Poitou-Charente).
Génération 2010 à 3 ans	33 500	Disposer d'environ 20 000 répondants pour les exploitations nationales du CEREQ et 13 000 répondants pour les extensions partenariales ciblées, qui couvrent aussi des sortants "post-initiaux" (sortants de l'enseignement supérieur, des formations sanitaires et sociales, des formations Jeunesse et Sports, des formations de l'environnement, des écoles de la fonction publique et des sortants des régions Normandie, Picardie et Provence-Alpes-Côte d'Azur).
Génération 2010 à 5 ans	13 700	La 2e interrogation a concerné les répondants au questionnaire de la 1ère interrogation destinés aux exploitations nationales du CEREQ.
Génération 2010 à 7 ans	8 900	La 3e interrogation a concerné les répondants de la 2e interrogation.
Génération 2013 à 3 ans	19 500	Disposer d'environ 10 000 répondants pour les exploitations nationales du CEREQ et 12 700 répondants pour les extensions partenariales ciblées, dont certaines couvrent aussi des sortants "post-initiaux" (sortants de l'enseignement supérieur, des formations sanitaires et sociales, des formations Jeunesse et Sports, des formations de l'environnement, sortants des zones urbaines sensibles).
Génération 2017 à 3 ans	25 200	Disposer d'environ 20 000 répondants pour les exploitations nationales du CEREQ et 16 000 répondants pour les extensions partenariales ciblées, dont certaines couvrent aussi des sortants "post-initiaux" (sortants de l'enseignement supérieur, des formations Jeunesse et Sports, des formations de l'environnement, des formations professionnelles de l'automobile, sortants des quartiers prioritaires de la politique de la ville et sortants des établissements des régions Bretagne et Pays de la Loire).

Notes

[1] La rénovation du dispositif « Génération » a été présentée au Conseil national de l'information statistique le 9 novembre 2017, dans une réunion consacrée à l'insertion professionnelle des jeunes. https://www.cnis.fr/wp-content/uploads/2017/09/CR_2017_2e_reunion_COM_EMPLOI-1.pdf

[2] Les extensions d'échantillon, dont les extensions régionales, sont présentées dans le tableau en annexe.

[3] Le tirage des stages à décrire est effectué automatiquement lors de la passation du questionnaire à partir de la liste des stages effectués et de leurs caractéristiques (durée, secteur, nombre), de façon à décrire au plus, deux stages par personnes. Rouaud (2012) présente la méthode d'échantillonnage des stages et leur pondération.

[4] D'après le dossier transmis au comité du label en vue de la labellisation « statistique publique » de l'enquête.

[5] La documentation du Cereq recense par exemple, au 15 décembre 2023, 61 références de publications, articles ou chapitres d'ouvrages, mobilisant l'enquête « Génération 2013 ». Pour l'enquête « Génération 1998 », la plus ancienne, qui a fait l'objet de quatre interrogations, c'est plus de 500 références.