



**HAL**  
open science

## Enquête Génération 2010 1 ère interrogation printemps 2013

Christophe Barret, Christophe Dzikowski, Zora Mazari, Pascale Rouaud,  
Florence Ryk, Noémie Volle

► **To cite this version:**

Christophe Barret, Christophe Dzikowski, Zora Mazari, Pascale Rouaud, Florence Ryk, et al.. Enquête Génération 2010 1 ère interrogation printemps 2013. [Rapport Technique] Céreq Etudes n°25, Céreq. 2019, 151 p. halshs-02889467

**HAL Id: halshs-02889467**

**<https://shs.hal.science/halshs-02889467>**

Submitted on 3 Jul 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# CÉREQ ÉTUDES

25  
2019

## Enquête Génération 2010 1<sup>ère</sup> interrogation printemps 2013

Bilan méthodologique

Christophe BARRET, Christophe DZIKOWSKI, Zora MAZARI,  
Pascale ROUAUD, Florence RYK, Noémie VOLLE

*Équipe ingénierie et gestion d'enquête,  
Département des Entrées et Évolutions dans la Vie Active (DEEVA), Céreq.*



# **Enquête Génération 2010**

## **Bilan méthodologique**

### **1<sup>ère</sup> interrogation – printemps 2013**

#### Auteurs

Christophe Barret  
Christophe Dzikowski  
Zora Mazari  
Pascale Rouaud  
Florence Ryk  
Noémie Volle

*Membres ou ex-membres de l'équipe ingénierie et gestion d'enquête du Département des Entrées et des Evolutions dans la Vie Active (DEEVA) du Céreq*

## Synthèse

Ce document de travail présente un panorama complet des travaux liés à l'enquête 2013 auprès de la Génération 2010. Après une brève présentation du dispositif des enquêtes Génération et plus spécifiquement de l'enquête 2013, sont détaillées toutes les étapes de réalisation de cette enquête dans un ordre chronologique. La première étape est la constitution de la base de sondage d'élèves présumés sortants du système éducatif en 2009/2010. Il s'agit d'une base de données d'élèves collectée auprès des établissements de formation initiale situés en France. Après avoir défini un plan de sondage stratifié et équilibré, le tirage aléatoire de l'échantillon est effectué à l'aide de l'algorithme du Cube. Après une étape de préparation (développement du Cati, enrichissement des coordonnées, lettre-avis, etc.), la collecte par téléphone s'est déroulée d'avril à juillet 2013. Enfin, l'ensemble des traitements post-collecte sont décrits (création des bases de données, apurement, codification, traitement de la non-réponse et calage, etc.).

## Sommaire

<b>I. Présentation de l'enquête 2013 auprès de la Génération 2010.....</b>	<b>6</b>
<b>1. Le dispositif des enquêtes Génération.....</b>	<b>6</b>
<b>2. Le champ de l'enquête .....</b>	<b>8</b>
<b>3. Les extensions.....</b>	<b>9</b>
<b>4. Le questionnaire .....</b>	<b>10</b>
a) Description du questionnaire.....	10
b) Zoom sur le calendrier professionnel .....	11
c) Les extensions de questionnement .....	13
<b>5. L'intérêt et les atouts du dispositif .....</b>	<b>14</b>
<b>6. Le calendrier de l'enquête .....</b>	<b>15</b>
<b>II. La constitution de la base de sondage.....</b>	<b>16</b>
<b>1. La collecte auprès des établissements.....</b>	<b>16</b>
a) Collecte des bases rectorales.....	16
b) Collecte des fichiers des diplômés du sport et de l'animation .....	17
c) Collecte des fichiers des universités .....	17
d) Collecte des fichiers des écoles de la santé et du social.....	17
e) Collecte des fichiers CIFRE et Allocataires .....	17
f) Collecte des fichiers auprès des DRJSCS .....	17
g) Collecte des fichiers auprès des CFA.....	18
h) Collecte des fichiers auprès des écoles de la fonction publique (DGAFP) .....	18
i) Collecte auprès des autres établissements.....	18
<b>2. La compilation et l'apurement de la base de sondage.....</b>	<b>21</b>
a) Les fichiers reçus.....	21
b) Les couplages et apurements des fichiers .....	22
c) Amélioration de la qualité de la base de sondage : les numéros de téléphones, les mails .....	31
d) Le géocodage de la base de sondage.....	33
<b>III. Le plan de sondage et la constitution de l'échantillon.....</b>	<b>34</b>
<b>1. Objectifs et difficultés du plan de sondage .....</b>	<b>34</b>
<b>2. Description du plan de sondage .....</b>	<b>34</b>
a) Travail préparatoire .....	34
b) Simulation de 20 000 questionnaires « représentatifs » Céreq.....	40
c) Simulation de 25 000 questionnaires Céreq. ....	40
d) Calcul des suppléments de tirage pour les extensions et coefficient de tirage final .....	41
e) Tirage de l'échantillon Global (y compris réserve).....	42
f) Tirage de l'échantillon principal.....	42
g) Calcul du taux de sondage .....	42
<b>3. Bilan et perspectives.....</b>	<b>43</b>
a) Petit historique des plans de sondage des enquêtes « Génération ».....	43
b) L'utilisation de la macro de tirage CUBE appliquée à notre plan de sondage .....	44
<b>IV. Préparation de la collecte.....</b>	<b>45</b>
<b>1. Développement du cati (enquête téléphonique).....</b>	<b>45</b>
a) Logiciel utilisé.....	45
b) Technique de développement du calendrier .....	45
c) Test du cati et calendrier associé .....	45
<b>2. Restructuration Normalisation Validation Postale des adresses, et rachat des déménagés ....</b>	<b>46</b>
<b>3. Location de mails auprès de la poste .....</b>	<b>48</b>

<b>4. Premier enrichissement des coordonnées téléphoniques.....</b>	<b>48</b>
c) Le protocole de recherche .....	48
d) Bilan de l'enrichissement .....	50
<b>5. Ordonnancement des numéros de téléphones .....</b>	<b>52</b>
<b>6. La lettre avis .....</b>	<b>52</b>
<b>7. Le cawi de préfidélisation et le numéro vert .....</b>	<b>53</b>
<b>8. Une recherche supplémentaire de numéros de téléphone en cours de plateau .....</b>	<b>55</b>
<b>9. La phase de pré qualification.....</b>	<b>55</b>
<b>10. Relations et contacts avec les enquêtés .....</b>	<b>55</b>
a) Site internet du Céreq .....	55
b) Réseaux sociaux : Facebook et Twitter .....	55
c) Mailings .....	57
<b>V. La collecte par téléphone.....</b>	<b>59</b>
<b>1. Calendrier et organisation générale de la collecte.....</b>	<b>59</b>
<b>2. Le suivi de la collecte en chiffres .....</b>	<b>60</b>
<b>3. Les règles de rappel .....</b>	<b>61</b>
<b>4. Des messages laissés sur répondeur .....</b>	<b>62</b>
<b>5. Suivi technique et personnes « qualité » .....</b>	<b>62</b>
<b>6. Les résidents étrangers à la date d'enquête .....</b>	<b>62</b>
<b>VI. Les traitements en aval .....</b>	<b>63</b>
<b>1. Les tables brutes.....</b>	<b>63</b>
<b>2. La création des tables .....</b>	<b>64</b>
a) La construction des bases brutes .....	64
b) Les variables de la table « Individus » construites à partir des calendriers .....	64
c) La vérification des filtres .....	65
d) La suppression d'observations non exploitables .....	65
e) Enrichissement par les informations de la base de sondage .....	66
f) Les recopies.....	66
g) L'anonymisation des fichiers.....	66
<b>3. La codification des variables.....</b>	<b>67</b>
a) Le repérage du plus haut diplôme obtenu .....	67
b) La codification des diplômes et spécialités de formation .....	67
c) La codification du secteur d'activité de l'établissement employeur .....	69
d) La codification de la profession des périodes d'emploi .....	71
e) La codification des contrats de travail .....	73
f) Les variables de localisation géographique.....	74
g) La codification des réponses en clair .....	74
<b>4. Le traitement des salaires (primes incluses).....</b>	<b>74</b>
<b>5. Les tables finales .....</b>	<b>80</b>
<b>VII. La pondération finale .....</b>	<b>82</b>
<b>1. Le principe général .....</b>	<b>82</b>
<b>2. Le poids de couverture.....</b>	<b>84</b>
<b>3. Le poids d'échantillonnage.....</b>	<b>85</b>
<b>4. Le poids relatif au fait de contacter l'individu ou un proche .....</b>	<b>86</b>
a) Modélisation de la probabilité de contact .....	86

b)	Cohérence de la modélisation .....	86
c)	Mise hors champ des individus qui résident à l'étranger .....	87
d)	Poids de contact des individus ayant complété un questionnaire dans le champ du Céreq .....	87
<b>5.</b>	<b>Le poids relatif au fait d'accepter de répondre .....</b>	<b>88</b>
a)	Modélisation de la probabilité d'accepter de répondre .....	88
b)	Cohérence de la modélisation .....	88
c)	Mise hors champ des individus à l'issue du questionnaire filtre .....	88
d)	Poids relatif au fait d'accepter de répondre pour les individus ayant complété un questionnaire dans le champ du Céreq .....	89
<b>6.</b>	<b>Probabilité de répondre à l'intégralité du questionnaire sachant que l'on appartient au champ de l'enquête.....</b>	<b>89</b>
a)	Modélisation de la probabilité de terminer le questionnaire .....	89
b)	Cohérence de la modélisation .....	90
c)	Poids relatif au fait de terminer un questionnaire exploitable pour les individus mis à disposition dans la table d'exploitation .....	91
<b>7.</b>	<b>Le calage sur marges .....</b>	<b>92</b>
<b>8.</b>	<b>Le récapitulatif .....</b>	<b>95</b>
<b>9.</b>	<b>La pondération des extensions .....</b>	<b>97</b>
<b>VIII.</b>	<b><i>Une estimation de la variance par calcul direct .....</i></b>	<b>98</b>
<b>1.</b>	<b>Estimation de la variance due à l'échantillonnage équilibré .....</b>	<b>98</b>
<b>2.</b>	<b>Estimation de la variance en prenant en compte le processus de non-réponse .....</b>	<b>99</b>
<b>3.</b>	<b>Estimation de la variance de l'estimateur à l'issue de la phase de calage .....</b>	<b>99</b>
<b>4.</b>	<b>Calcul de la variable linéarisée .....</b>	<b>100</b>
<b>5.</b>	<b>Exemple de variance obtenue .....</b>	<b>101</b>
<b>IX.</b>	<b><i>Publications récentes.....</i></b>	<b>102</b>
<b>X.</b>	<b><i>Perspectives pour les enquêtes futures.....</i></b>	<b>103</b>
<b>XI.</b>	<b><i>Annexes .....</i></b>	<b>104</b>
	<b>Annexe 1 – Table des illustrations .....</b>	<b>104</b>
	<b>Annexe 2 – Lettres avis de contact avec les jeunes .....</b>	<b>106</b>
	Lettre envoyée par courrier POSTAL .....	106
	Lettre envoyée par courrier ELECTRONIQUE.....	110
	Organisation de l'envoi des lettres avis.....	112
	<b>Annexe 3 – Nomenclature des diplômes .....</b>	<b>113</b>
	<b>Annexe 4 – Nomenclature des niveaux d'études .....</b>	<b>117</b>
	<b>Annexe 5 – Nomenclature des spécialités.....</b>	<b>122</b>
	<b>Annexe 6 – Nomenclature NAF rev2 en 88 divisions (A88) .....</b>	<b>125</b>
	<b>ANNEXE 7 – Procédure de codification de l'activité selon la NAF rev2 .....</b>	<b>127</b>
	<b>Annexe 8 – Définition des variables annexes utilisées dans Sicore PCS .....</b>	<b>131</b>
	<b>Annexe 9 – Liste des questions ouvertes .....</b>	<b>135</b>
	<b>ANNEXE 10 – Pondérations des extensions.....</b>	<b>136</b>
a)	Extension DREES (formations sanitaires et sociales) (PB NUM) .....	136
b)	Extension SPORT (formations sport et animations).....	140
c)	Extension DGAFP (formations de la fonction publique) .....	143

## *1. Présentation de l'enquête 2013 auprès de la Génération 2010*

### **1. Le dispositif des enquêtes Génération**

Les enquêtes Génération s'intéressent à l'insertion et au cheminement des sortants du système éducatif lors de leurs premières années de vie active. Elles ont pour objectifs principaux de produire des indicateurs d'insertion (taux d'emploi, taux de chômage, taux d'emploi à durée indéterminée, etc.), selon les niveaux de formation, les filières, les spécialités, à destination des acteurs publics et sociaux. Elles proposent ainsi des informations qui contribuent à la compréhension des processus d'insertion et des différenciations des parcours en début de carrière.

#### **Encadré 1 - Le Centre d'Etudes et de Recherches sur les Qualifications**

Créé au début des années 1970, devenu établissement public en 1985, le Centre d'études et de recherches sur les qualifications (Céreq) est placé sous la tutelle des ministères en charge de l'éducation nationale et du travail.

Il a pour missions de développer des études et des recherches, de collecter et d'exploiter des données originales dans le domaine de la relation formation-emploi, et de formuler des avis et propositions destinés à éclairer les choix en matière de politiques de formation.

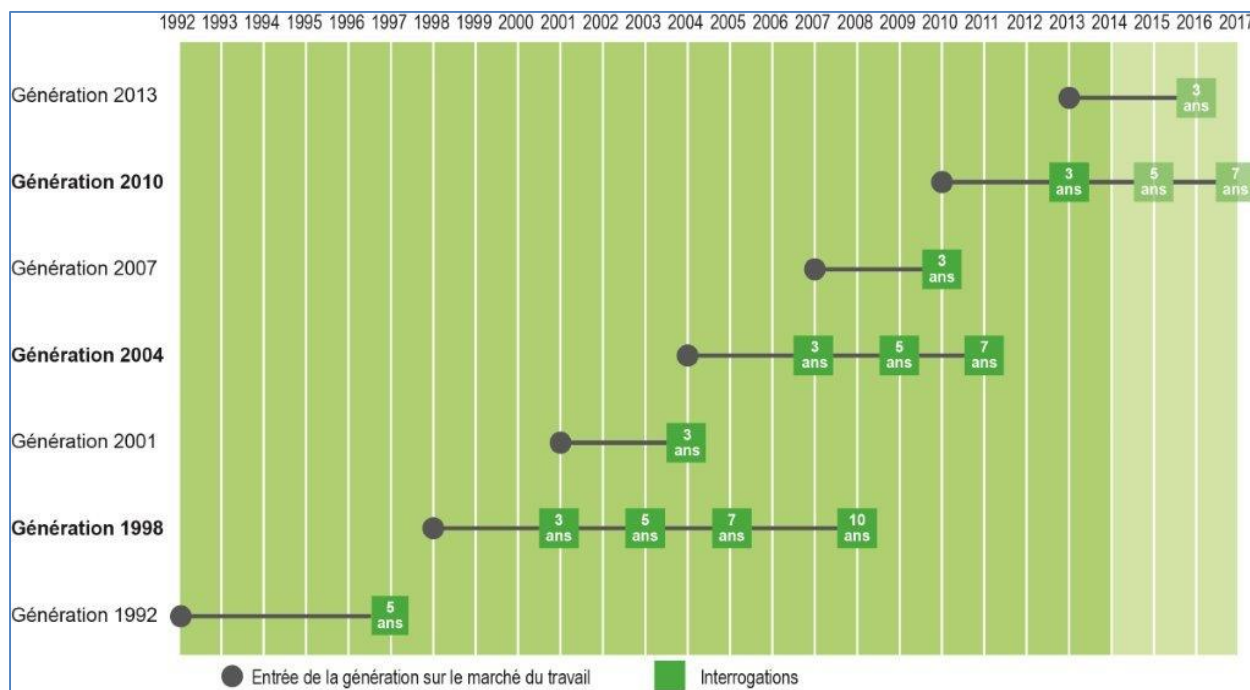
La première enquête « Génération » a été effectuée en 1997 auprès de jeunes sortis du système éducatif en 1992 et issus de tous les niveaux de formation. Un dispositif régulier d'interrogation a été mis en place à partir de l'enquête effectuée en 2001 auprès de sortants du système éducatif en 1998: les enquêtes s'enchaînent au rythme d'une génération nouvelle de sortants interrogée tous les trois ans, et avec une alternance entre une enquête « génération pleine » et une « mini-génération ».

L'enquête « génération pleine » comprend plusieurs interrogations. La première interrogation, trois ans après la sortie du système éducatif, est principalement tournée vers la production d'indicateurs et des usages institutionnels.

Les interrogations suivantes, à cinq et sept ans, sont plutôt centrées sur les usages analytiques (notamment sur la question des parcours, des mobilités sur moyen terme). Ces ré-interrogations permettent aussi d'approfondir certains constats issus de l'exploitation de la première interrogation (par exemple sur les insertions les plus problématiques).

L'enquête « mini-génération » se limite à la première interrogation à trois ans, avec une taille d'échantillon plus réduite, et un questionnement allégé. Cette enquête a vocation à réactualiser les indicateurs d'insertion sur les trois premières années d'insertion selon une grille d'analyse moins fine que pour les « générations pleines ».



**Figure 1 - Calendrier des enquêtes Génération**

Le cœur de l'enquête, pour la première interrogation comme pour les autres quand il y a lieu, est constitué par le calendrier professionnel qui permet de suivre mois par mois la situation des jeunes à l'issue de leur formation initiale, et de décrire les séquences successives d'emploi et de non emploi. Les enquêtes « Génération » permettent aussi de répondre à des demandes d'extensions nationales ou régionales portant sur les jeunes issus de certains niveaux ou spécialités de formations ou sur les jeunes ayant bénéficié de certaines mesures pour la formation. Les échantillons peuvent donc être élargis en conséquence.

**Tableau 1 - Détail des enquêtes Génération et effectifs de répondants**

	Année de la première interrogation	Nombre de répondants	Dont liés aux extensions d'échantillon
Génération 1992	1997	26 000	
Génération 1998	2001	55 000	19 000
Génération 2001	2004	25 000	15 000
Génération 2004	2007	65 000	37 000
Génération 2007	2010	29 000	26 000
Génération 2010	2013	38 500	13 500

## 2. Le champ de l'enquête

La Génération 2010 concerne les « primo sortants » de formation initiale en 2009-2010 (année scolaire). Cette restriction aux primo-sortants signifie que les sortants de formation qui avaient déjà interrompu leurs études pour une durée supérieure à un an avant l'année scolaire considérée sont hors champ. Tous les niveaux et domaines de formations sont concernés par l'enquête. De façon plus précise, les critères d'éligibilité pour être dans le champ retenu, nommé ensuite « champ Céreq », sont les suivants :

- avoir été inscrit dans un établissement de formation en France (Métropolitaine + DOM) durant l'année scolaire 2009-2010 ;
- avoir quitté le système éducatif entre octobre 2009 et octobre 2010<sup>1</sup> ;
- ne pas avoir interrompu ses études durant une année ou plus avant l'année scolaire 2009-2010 (sauf pour raison de santé) ;
- ne pas avoir repris ses études pendant l'année qui a suivi l'entrée sur le marché du travail ;
- avoir 35 ans ou moins en 2010 ;
- être localisé en France (Métropolitaine + DOM) au moment de l'enquête (ce qui exclut donc notamment les personnes résidant à l'étranger à la date d'enquête).

Quelques points particuliers concernent l'application de ces critères :

- bien qu'il s'agisse de contrats de travail au même titre que les contrats de professionnalisation ou que les contrats de qualification, les contrats d'apprentissage sont considérés comme relevant de la formation initiale. Une personne repérée comme sortant de formation en 2009-2010 qui poursuit par un contrat d'apprentissage en 2010-2011 est donc considérée en poursuite d'études, donc hors champ. Ce n'est pas le cas s'il s'agit d'un contrat de professionnalisation ;
- une personne sortie d'un établissement de formation en 2009-2010 qui poursuit des cours par correspondance ou des cours du soir en 2010-2011, est considérée comme en poursuite d'études, donc hors champ, si elle n'a pas d'emploi en parallèle. Si elle a un emploi en parallèle, sa situation d'emploi prime ;
- une personne sortie d'un établissement de formation en 2009-2010 mais poursuivant ses études en 2010-2011 comme élève fonctionnaire n'est pas considérée comme étant en poursuite d'études ; elle est considérée en emploi.

Les différences de champ qui existent entre les quatre cohortes enquêtées sont mineures :

- les sortants de classes de 6<sup>e</sup> et 5<sup>e</sup>, inclus dans le champ de la Génération 1998 et de la Génération 2001 sont exclus du champ depuis la Génération 2004 ;
- les étudiants étrangers sortants de l'Université (repérables dans le Système d'Information sur le Suivi de l'Étudiant, SISE) étaient hors champ pour la Génération 1998 et la Génération 2001; ils sont inclus dans le champ depuis la Génération 2004 ;
- la période de référence pour la date de sortie était l'année civile pour la Génération 1998 et la Génération 2001 ; c'est l'année scolaire depuis la Génération 2004 (octobre n à octobre n+1) ;

---

<sup>1</sup> Pour certaines formations (école fonction publique, sport/animation, santé/social, et thèse) il y a une prolongation jusqu'à décembre 2010

- la réforme de mastérisation des concours d'enseignement, mise en place sous régime transitoire l'année 2010, et surtout le rattachement administratif des IUFM aux universités, n'ont pas permis d'intégrer dans cette enquête Génération les sortants de 1<sup>re</sup> année de cette formation.

### Encadré 2 - Une enquête de la statistique publique

L'enquête 2013 auprès de la Génération 2010, extensions comprises, est une enquête de la statistique publique et relève à ce titre de la loi n° 51-711 du 7 juin 1951 sur l'obligation, la coordination et le secret en matière de statistiques. Celle-ci définit le cadre de procédures destinées à garantir l'intérêt, la pertinence et la qualité des enquêtes publiques, ainsi que la confidentialité des informations collectées. L'opportunité de l'enquête 2013 auprès de la Génération 2010 a ainsi été discutée au Comité national de l'information statistique (CNIS), qui a émis un avis favorable en mars 2012 puis lui a attribué un label d'intérêt général et de qualité statistique en octobre 2012 (label n° 2013X705AU). Les réponses apportées à l'enquête sont confidentielles. Les fichiers des réponses détaillées à l'enquête qui sont mis à disposition des chercheurs sont anonymes et ne comportent pas d'informations susceptibles de permettre une identification directe ou indirecte des personnes enquêtées. La réalisation de l'enquête et sa diffusion a fait l'objet d'une déclaration auprès de la Commission nationale de l'informatique et des libertés (CNIL).

## 3. Les extensions

Les acteurs intervenant dans le domaine de la formation sont multiples, avec leurs questionnements propres mais aussi avec un questionnement commun sur le devenir des bénéficiaires. A la demande de certains d'entre eux, ministères, conseils régionaux ou observatoires de branches, des extensions sont régulièrement adossées aux enquêtes Génération. Elles permettent de situer les analyses ciblées sur des publics, des filières ou des territoires particuliers dans un contexte plus large, par rapprochement avec des indicateurs de référence nationaux.

Plusieurs types d'extensions existent, parfois combinées :

- l'extension d'échantillon, pour disposer d'un nombre de répondants suffisant pour permettre des analyses statistiquement pertinentes sur les catégories de sortants ciblées ;
- l'extension de champ, pour interroger également sur les catégories de sortants ciblées les personnes non retenues dans le champ Céreq (par exemple en levant la restriction de primo sortie) ;
- l'extension de questionnement, pour poser quelques questions complémentaires sur des catégories de sortants ciblées.

Le tableau suivant présente pour chaque partenariat mené dans le cadre de l'enquête Génération, le nombre de questionnaires conventionnés et attendus ainsi que le type d'extension.

**Tableau 2- extension de l'enquête Génération**

Institution ou organisme demandeur	Population cible	Champ	Nombre de questionnaires conventionnés	Nombre de questionnaires exploitables	type d'extension
Région Picardie	Sortants de l'enseignement secondaire en Picardie	Céreq	2000	2126	échantillon
DREES	Sortants de formations sanitaires et sociales	Céreq + post-initiaux	5000 (3400 formations santé, 1600 formations sociales)	6459 (4262 formations santé, 2197 formations sociales)	champ
CGDD	Sortants de formations environnementales	Céreq	3000	2724	questionnement et échantillon
Agefiph	Tous les répondants dans le champ Céreq	Céreq		33547	questionnement
Région PACA	Sortants de l'enseignement secondaire en PACA	Céreq	1500	1833	échantillon
Région Nord-Pas-de-Calais	Sortants formés en Nord-Pas-de-Calais	Céreq	1500	2975	échantillon
Ministère Jeunesse et Sport	Sortants des formations du sport	Céreq + post-initiaux	1500 (avenant + 240)	1707	champ
DGESIP - Thèse	Sortants diplômés de thèse	Céreq	1500	1911	questionnement et échantillon
DGESIP - Autres sortants	Sortants du supérieur	Céreq	2400 (800 sans diplôme du supérieur, 800 diplômés de licence générale, 800 diplômés de M1)	2054 1838 582	questionnement et échantillon
DGESIP - Module Professionnalisation	Sortants du supérieur en année diplômante, hors thèses	Céreq	14000	14454	questionnement et échantillon
DGAFP - Sortants de la fonction publique	Sortants d'une école de la fonction publique	Céreq + post-initiaux + écoles hors du champ du Céreq	1250	1643	champ
DGAFP - Module stage	Tous les répondants dans le champ Céreq, hors apprentis	Céreq	11 200 sortants n'ayant pas effectué de stage / 16 800 ayant effectué un stage	9 978 sortants n'ayant pas effectué de stage / 18 169 ayant effectué un stage	questionnement

## 4. Le questionnaire

### a) Description du questionnaire

Cette enquête est la première interrogation trois ans après leur sortie du système éducatif, des sortants de formation initiale de l'année scolaire 2009-2010.

Le questionnaire commence par une partie filtre destinée à valider l'identification de l'individu et à vérifier les critères d'éligibilité : l'interlocuteur est-il le « bon individu » ? Fait-il partie du champ (champ Céreq, champ d'une extension ou hors champ) ?

Le questionnaire de base aborde ensuite successivement les thèmes suivants : le parcours scolaire, l'identification du plus haut diplôme, les diplômes obtenus après 2010, le calendrier mensuel d'activité sur les trois années suivant la sortie du système éducatif avec une description détaillée de l'ensemble des séquences, les caractéristiques individuelles et l'environnement familial.

A ce questionnaire de base s'ajoutent des modules de questionnement développés en interne (Céreq) pour répondre à des besoins de partenaires d'extension ou des besoins de recherche :

- séjours à l'étranger ;
- stages en cours d'études ;
- professionnalisation dans l'enseignement supérieur (nouveau module) ;

- expériences de travail en cours d'études ;
- module handicap et problème de santé durable.

Trois extensions à cette enquête comportent un élargissement du champ Céreq portant sur les sortants :

- des formations « Sport » ;
- des formations « Santé-social » ;
- des écoles de la fonction publique.

Ces échantillons concernent des primo-sortants et des post-initiaux (ayant déjà arrêté plus d'un an leurs études auparavant). Ces extensions d'échantillon n'ont aucune incidence sur l'articulation du questionnaire.

Le nombre de répondants à l'enquête 2013 auprès de la Génération 2010 est de 38 594 individus dont :

- 33 547 questionnaires dans le champ Céreq (dont environ 8 500 issus d'extension dans le champ Céreq) ;
- 1 198 questionnaires supplémentaires pour l'extension des écoles de la fonction publique (DGAFP) ;
- 2 677 questionnaires supplémentaires pour l'extension « Santé-social » ;
- 1 172 questionnaires supplémentaires pour l'extension « Sport-Animation ».

Le plateau de l'enquête 2013 auprès de la Génération 2010 s'est déroulé entre le 2 avril et le 31 juillet 2013. Le questionnaire d'enquête a eu une durée moyenne d'environ 34 minutes.

## **b) Zoom sur le calendrier professionnel**

Le calendrier professionnel est le cœur du questionnaire. Il retrace mois par mois le parcours de l'enquêté entre sa date de fin d'études et la date de l'enquête, et pilote l'ouverture des différents modules qui permettront un descriptif détaillé des séquences qui y sont repérées.

Le calendrier permet de distinguer six situations exclusives l'une de l'autre :

- situation d'emploi en entreprise ;
- situation d'emploi en intérim
- situation de recherche d'emploi ;
- situation de reprise d'études ;
- situation de formation<sup>2</sup> ;
- autres situations.

---

<sup>2</sup> Les situations de contrat d'apprentissage (plus d'un an après la sortie), les situations de contrat de professionnalisation ou d'élève fonctionnaire sont considérées comme des périodes d'emploi « en entreprise » et traitées comme telles.



Pour les séquences passées, il est demandé en plus le motif du départ de l'entreprise ; pour les séquences actuelles (à la date d'enquête), l'opinion sur la situation d'emploi.

Pour les séquences de non emploi (recherche d'emploi, formation, autres) :

Rôle de Pôle Emploi (inscription, entretien(s) avec un conseiller, utilité, offres d'emploi proposées), démarches de recherche d'emploi personnelles, formations suivies (objectifs et financement), aides financières perçues (indemnités chômage et allocations).

Pour la séquence actuelle, en plus, le montant de l'aide financière

Pour les séquences de reprise d'études :

But de la reprise d'étude et rôle du service public de l'emploi dans la reprise d'étude.

Pour la séquence actuelle, en plus, le diplôme ou le type de concours préparé

Remarque : Dans le questionnaire, les questions concernant les séquences du passé sont formulées au passé et correspondent aux variables commençant par **EP**. Celles concernant la séquence actuelle sont formulées au présent et correspondent aux variables commençant par **EA**. A noter que lorsque les questions sont communes aux séquences passées et à la séquence actuelle, les 2 questions **EA** et **EP** sont au final regroupées dans le fichier d'enquête dans une seule variable intitulée **EP**.

### c) Les extensions de questionnement

Les extensions de questionnement concernent des modules prévus avec nos partenaires d'extension.

#### **Module « Post-initiaux des formations du domaine Santé/Social »**

Ce module concerne uniquement les sortants de l'extension d'échantillon Santé/Social qui ne sont pas primo sortants. Pour eux, le module se substitue aux questions sur le parcours scolaire et les expériences de travail en cours d'études.

Les questions abordent l'existence d'une formation antérieure dans le domaine, la situation d'activité avant la dernière formation, les objectifs de cette dernière, l'origine de son financement, le plus haut diplôme obtenu, le baccalauréat obtenu.

#### **Module « Post-Initiaux des formations du sport et de l'animation »**

Ce module concerne uniquement les diplômés de l'extension d'échantillon sur les formations du sport et de l'animation qui ne sont pas primo sortants. Pour eux, le module se substitue aux questions sur le parcours scolaire.

Les questions abordent l'existence d'une formation antérieure dans le domaine du sport ou de l'animation, la situation d'activité avant la dernière formation, les objectifs de cette formation, l'origine de son financement, le plus haut diplôme obtenu, le baccalauréat obtenu, la détention de diplômes et brevets dans le domaine du sport et de l'animation.

#### **Module « thèse »**

Ce module concerne tous les sortants de thèse ayant réussi leur doctorat (exceptés ceux du domaine Santé). Lors de la passation du questionnaire, ce module intervient après les questions sur le parcours scolaire et les séjours à l'étranger dans le cadre de la scolarité.

Les questions du module abordent le projet professionnel au moment de la soutenance de thèse, le cadre institutionnel de la thèse (nature de la rémunération, nature du laboratoire), la publication de travaux avant la soutenance, la situation post doctorale (stages post doctorat, qualification au CNU).

### **Module « Santé Handicap »**

Ce module concerne tous les répondants, qu'ils soient dans une extension d'échantillon ou non. Il est passé en fin de questionnaire, avant la partie consacrée aux discriminations ressenties.

Plusieurs questions abordent successivement l'existence d'un problème de santé durable ou un handicap ainsi que sa nature, à quel moment c'est apparu, si c'est un handicap reconnu et quelle est la dénomination de cette reconnaissance, et pour finir si ce problème de santé ou handicap a été déclaré aux cours de la recherche d'emploi.

### **Module Développement durable**

Ce module concerne les sortants des formations de développement durable. Quatre questions abordent successivement la motivation dans le choix de ce type de formation et ce qu'ils estiment avoir comme possibilités de débouchés, leur avis sur la dégradation de l'environnement, et pour finir sur leur avis sur l'avenir des métiers verts.

## **5. L'intérêt et les atouts du dispositif**

### **Un cadre d'analyse homogène et cohérent**

Contrairement à d'autres enquêtes d'insertion qui visent des publics segmentés (apprentis, lycéens, sortants de grandes écoles ou d'université...), seul le dispositif « Génération » propose un questionnement, une méthodologie et un cadre d'analyse homogène pour tous, quels que soient le parcours scolaire, les diplômes obtenus, les domaines et voies de formation. Il est donc possible de comparer et d'évaluer l'impact de ces différentes caractéristiques sur les variations observées au cours des premières années de vie active : qui accède rapidement à un emploi ? Qui reste durablement au chômage ? A quel type d'emploi accède-t-on ? A quel niveau de rémunération ? Telles sont les questions auxquelles le dispositif permet de répondre. Plus généralement, il met en évidence les phénomènes de concurrence ou de complémentarité entre niveaux, domaines et voies de formation.

### **Des informations riches et diversifiées**

Grâce à un questionnaire détaillé et un échantillon important, les enquêtes permettent, au-delà des caractéristiques du parcours scolaire et des diplômes obtenus, de prendre en compte d'autres critères. Le genre, l'origine sociale, l'origine nationale, le lieu de résidence, les mobilités géographiques, le statut familial, les réseaux sociaux mais aussi la place et le rôle des dispositifs publics sont autant de dimensions que le dispositif permet d'intégrer pour analyser les différences observées au cours des premières années de vie active.

### **Un recul temporel nécessaire**

Certaines enquêtes d'insertion sont réalisées quelques mois seulement après la sortie du système scolaire. L'option retenue est alors de disposer d'indicateurs qui peuvent être mis rapidement à disposition des décideurs, des familles et des étudiants. Avec le dispositif « Génération », la première interrogation est réalisée trois ans après la sortie du système scolaire. Les résultats des premières enquêtes ont mis en évidence l'importance de ce recul temporel. En effet, il faut attendre plusieurs années pour que la stabilisation professionnelle soit établie pour le plus grand nombre. Enquêter tôt après la sortie de formation donne une photographie faussée des situations par rapport à l'emploi, qui accentue fortement les différences, alors que les enquêtes « Génération » montrent que celles-ci tendent à se réduire avec le temps.



## Un suivi longitudinal

Le questionnaire permet aux jeunes débutants de décrire systématiquement, mois par mois, les différentes situations qu'ils ont connues depuis leur sortie du système éducatif. Ce mode d'interrogation permet de construire différents indicateurs comme le taux de chômage ou le taux d'emploi, et d'aborder la qualité de l'emploi (niveau de rémunération, type de contrat). Il permet aussi de construire des typologies de parcours à partir de la description des situations mois par mois. Ces typologies offrent une vision synthétique des premières années sur le marché du travail : trajectoire d'accès rapide à l'emploi, trajectoire d'accès différé à l'emploi, trajectoire de décrochage, etc. L'insertion est une réalité multidimensionnelle qui ne peut se réduire à un ou deux indicateurs.

## La même conjoncture pour tous

Les « générations » sont construites en fonction de la date de sortie de formation et non de l'année de naissance. Quel que soit leur niveau de formation, les jeunes arrivent donc dans un contexte de marché du travail plus ou moins favorable mais identique pour tous. Il est donc plus facile a priori de comparer les trajectoires d'accès à l'emploi. Mais cette conjoncture a-t-elle les mêmes effets pour tous : à qui profitent les embellies ? Qui souffrent le plus des retournements ? Quels effets sur les taux de chômage, l'importance des CDD ou de l'intérim, et pour qui ? Telles sont les questions auxquelles le caractère récurrent des enquêtes « Génération » permet de répondre.

## 6. Le calendrier de l'enquête

La réalisation de l'enquête s'est étendue sur quatre années. Les bases de sortants de formation initiale ont été collectées dès leur disponibilité, au cours de l'année 2011. Le questionnaire a été élaboré au cours de l'année 2012 et du premier trimestre 2013. Les extensions ont été définies sur la même période.

L'enquête a été réalisée par téléphone lors du 2<sup>d</sup> trimestre 2013, la phase de test de l'enquête commençant dès le 1<sup>er</sup> trimestre 2013.

La chronologie des principales étapes de l'enquête « Génération 2010 » a été la suivante :

Etapes	Début	Fin
1. Collecte des bases établissements*	Janvier 2012	Mai 2012
2. Conception et test papier des modules de questionnement	Juin 2012	Juillet 2012
3. Elaboration des conventions	Septembre 2012	Novembre 2012
6. Tirage de l'échantillon	Septembre 2012	Décembre 2012
4. Conception du Cati	Novembre 2012	Avril 2013
5. Tests Cati (x 3)	11 février 2013	10 mars 2013
7. Recherches de coordonnées téléphoniques	15 janvier 2013	4 février 2013
8. Terrain d'enquête	3 Avril 2013	31 juillet 2014
9. Constitution des bases, apurements, codifications, pondération, rédaction, documentation		Juillet 2014
10. Livraison des bases aux extensions		22 juillet 2014
11. Bilan méthodologique	Septembre 2014	Octobre 2014

*\*cette période concerne la collecte des établissements hors rectorats (collecte qui a été réalisée en interne en mai 2011)*

## II. La constitution de la base de sondage

Il n'existe pas de base d'élèves centralisée couvrant l'ensemble des formations initiales dispensées en France. Il faut donc la constituer spécifiquement pour l'enquête, à partir de différentes sources, dans le cadre d'une collecte *ad hoc*.

La base ainsi constituée présente un défaut de couverture dû aux établissements qui n'ont pas fourni de bases d'élèves lors de la constitution de la base de sondage.

Par ailleurs, les données récupérées sont souvent des listes de sortants d'un établissement : il s'agit de personnes inscrites en 2009-2010 dans l'établissement et non réinscrites en 2010-2011 dans le même établissement. Elles peuvent néanmoins s'être réinscrites ensuite ailleurs et sont alors hors du champ de l'enquête. La base comporte aussi des doublons : les jeunes inscrits dans plusieurs établissements ou dans plusieurs filières qui ne seraient pas repérés.

Il n'est possible de traiter que partiellement ces deux aspects lors de la constitution de la base de sondage. Ils seront alors traités pendant l'enquête.

### 1. La collecte auprès des établissements

Pour chaque individu scolarisé pendant l'année scolaire 2009-2010 (présupposé sortant), les informations suivantes devaient être récupérées : nom et prénoms, adresse et numéros de téléphone, mail, date et lieu de naissance, sexe, diplôme et spécialité de formation, obtention ou non du diplôme préparé.

Des informations relatives à l'établissement de formation étaient également collectées. Lorsque le numéro d'identifiant national d'étudiant (INE) était disponible (informations issues des rectorats, des universités et des CFA), il était récupéré pour les traitements d'apurement de la base (repérage des poursuites d'études et des doublons). Dans ce cadre et pour les CFA, les fichiers d'inscrits en 2010-2011 ont également été récupérés.

#### a) Collecte des bases rectoriales

Les rectorats disposent des listes d'inscrits dans les établissements de leur ressort : collèges, lycées et BTS dépendant du ministère en charge de l'éducation nationale (y compris les établissements privés sous contrat), grâce au système informatique *Scolarité*.

Le Céreq a réalisé lui-même la collecte des informations souhaitées auprès des rectorats, initialisée par un mail de la DEPP, et selon un dessin d'extraction également défini avec la DEPP.

La collecte a eu lieu au second trimestre 2011. Elle a permis de récupérer pour chaque académie la liste de l'ensemble des inscrits au cours de l'année scolaire 2009-2010, ainsi que celle des inscrits en 2010-2011.

La couverture des établissements est ainsi exhaustive sur le champ rectoral. Au total, les établissements couverts par les rectorats représentent 82 % des 17 000 établissements de la base de sondage finale et concernent, d'après les résultats de l'enquête, 44 % des sortants<sup>5</sup>.

---

<sup>5</sup> Par ailleurs, les universités représentent 0,6 % des établissements de la base de sondage finale et 28 % des sortants ; les autres établissements, 17,4 % des établissements et 28 % des sortants.

## **b) Collecte des fichiers des diplômés du sport et de l'animation**

Pour la Génération 2007, cette collecte avait été réalisée en interne en collaboration avec le ministère de la Jeunesse et des sports pour lequel le Céreq réalise une extension d'échantillon et de questionnement.

Pour cette Génération 2010, les DRJSCS ont été intégrées à la collecte globale des établissements et n'ont pas été traitées indépendamment.

## **c) Collecte des fichiers des universités**

Contrairement à la génération précédente, la collecte des universités a été intégrée à la collecte générale. Pour la génération 2007, cette opération avait été traitée en interne, notamment à cause de la mise en place de l'autonomisation des universités, la loi LRU.

En cours de collecte, le réseau sup-cil (réseau des correspondants informatiques et libertés des universités) nous a signalé l'existence de la délibération RU13 du 10 décembre 2009, délibération Cnil, sur le traitement de la gestion de la scolarité des étudiants mis à disposition par l'Amue. Cette délibération ne mentionnait pas le Céreq en destinataire des données, données gérées par les logiciels « Apogee ».

Cette collecte a donc été décalée dans le temps (fin mars) afin d'être en règle avec la Cnil et avec la délibération RU13. Pour anticiper et pour accélérer le travail de chaque université (suite à la régularisation du dossier), il a été décidé avec l'Amue de développer un patch d'extraction des sortants universitaires.

## **d) Collecte des fichiers des écoles de la santé et du social**

Les écoles de la santé et du social ne sont pas extraites de la BCE. Notre demande est effectuée auprès de la Drees (pour récupérer les bases établissements issues de FINESS) car les bases réceptionnées nous permettent une sélection précise des formations. Les nomenclatures associées nous sont fournies au même moment.

Au cours de la collecte, une relance courrier a été effectuée directement par la Drees auprès de ces écoles afin d'optimiser la collecte des fichiers de ce champ.

## **e) Collecte des fichiers CIFRE et Allocataires**

Comme pour Génération 2007, la collecte des fichiers CIFRE a été réalisée en interne.

Cette collecte a été faite tardivement, en avril 2012. Une structure de fichier a été élaborée pour la première fois, en mentionnant la totalité des informations nécessaires à la création de la base des sortants de ce champ.

Le fichier centralisé complémentaire sur les allocataires de thèses n'étant plus collecté, ils ont été intégrés à la collecte des universités.

## **f) Collecte des fichiers auprès des DRJSCS**

La collecte des fichiers du sport autrefois réalisée en interne a été réintégrée à la collecte générale. Aucune convention de collecte n'a été établie car elle n'était pas justifiée.

Un mail du ministère de la Jeunesse a été envoyé à toutes les directions régionales pour soutenir notre demande et ainsi optimiser la collecte.

### **g) Collecte des fichiers auprès des CFA**

La plupart des CFA disposent d'une application de gestion de la scolarité. Pour la deuxième fois, en collaboration avec le développeur de l'application, un extract homogène a pu être développé et utilisé par une grande majorité de CFA. Cette opération a été d'une grande efficacité puisque les fichiers étaient disponibles facilement pour les CFA (presse bouton) et homogènes, ce qui a facilité le traitement.

En parallèle de cette collecte, les fichiers SIFA nationaux ont été collectés comme source de référence pour les deux années consécutives 2009-2010 et 2010-2011.

Ces fichiers nous ont permis notamment de récupérer des informations plus précises sur les diplômés et de supprimer les réinscrits l'année suivante.

### **h) Collecte des fichiers auprès des écoles de la fonction publique (DGAFP)**

La DGAFP a souhaité réaliser de manière expérimentale, une extension d'échantillon sur les sortants des écoles de la fonction publique pour la génération 2007. Les résultats ayant été concluants, cette extension a été renouvelée dans le cadre de la Génération 2010.

Les établissements de la fonction publique ne sont renseignés que pour moitié dans la BCE. Comme auparavant, une liste a donc été demandée à la DGAFP et retravaillée avec eux, en vue de les intégrer dans le système de collecte. Par ailleurs, un courrier de sensibilisation a été rédigé par la DGAFP pour appuyer notre demande, nous garantissant un taux de réponse satisfaisant.

### **i) Collecte auprès des autres établissements**

Selon la BCE, en dehors des établissements dépendants des rectorats, environ 4 100 établissements entrent dans le champ de l'enquête (dispensant de la formation initiale) : universités, « grandes écoles », lycées agricoles, centres de formation des apprentis ....

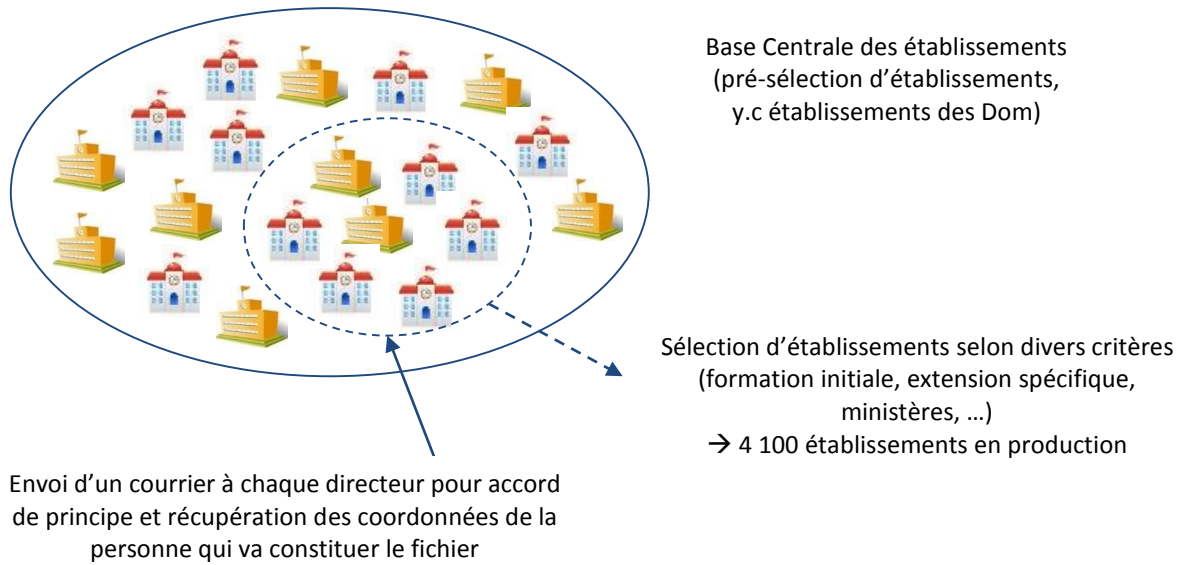
Le Céreq a confié à un prestataire ces contacts. Notre collecte de fichiers s'est déroulée de février à mai 2012 selon le schéma organisationnel de collecte suivant (Figure 3 page 19).

#### **Encadré 4 - La base centrale des établissements (BCE)**

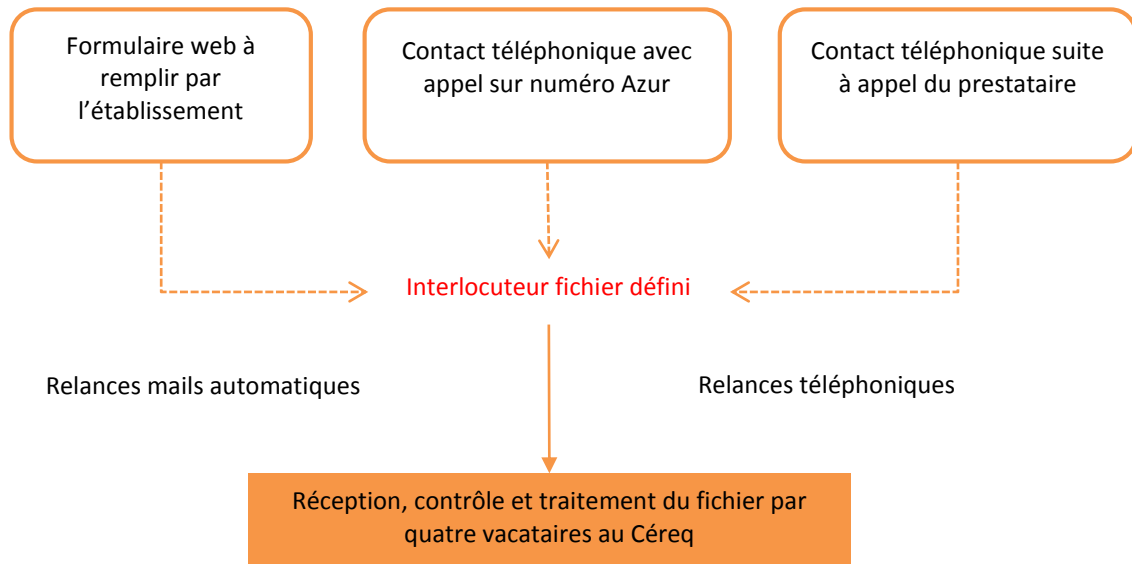
La BCE, gérée par le ministère en charge de l'éducation nationale, est le répertoire national des établissements assurant une activité de formation initiale générale, technique ou professionnelle, de la maternelle à l'enseignement supérieur, qu'ils soient publics ou privés, sous tutelle ou non du ministère de l'Education nationale et de la recherche. Elle concerne également les structures d'administration du système éducatif public ainsi que certaines parties d'établissements qui ont besoin d'être identifiées pour la gestion du système éducatif. La BCE a pour rôle d'alimenter toutes les applications informatiques nationales de gestion administrative, financière, statistique ou documentaire du ministère de l'Education nationale et de la recherche. Elle est mise à jour à partir des bases rectorales d'établissements, en temps réel.

Source : <http://www.infocentre.education.fr>

**Figure 3 - Schéma organisationnel de la collecte**



Collecte des coordonnées de la personne qui va constituer le fichier par  
différents modes de contact  
(à l'aide d'un prestataire)  
*(via à un prestataire de plateau téléphonique)*



Le déroulement de la collecte a été le suivant :

- *Phase 1 (prestataire)* : envoi d'un courrier signé par le directeur du Céreq sollicitant la mise à disposition des données individuelles nécessaires à la constitution de la base de sondage et d'un courrier technique à l'attention de la personne qui réalisera l'extraction du fichier. Les objectifs de ces courriers sont les suivants : l'obtention de l'accord de principe du responsable de l'établissement, la désignation d'un correspondant technique et la transmission à ce correspondant des instructions nécessaires à la constitution des fichiers. Ce courrier prévoyait la transmission du nom et des coordonnées du correspondant par questionnaire web ou par téléphone (n° Azur).
- *Phase 2 (prestataire)* : échanges téléphoniques avec les établissements et fixation d'une date de transmission de fichier.
- *Phase 3 (prestataire)* : relances téléphoniques et mail de l'interlocuteur en cas de non réception du fichier à la date convenue.

Hormis les contacts initiaux par courrier, les contacts avec les établissements ont été principalement téléphoniques assistés par un système Cati, puis par mail. Afin de diminuer les taux de refus des établissements, plusieurs ministères de tutelles des établissements, préalablement sensibilisés à l'enquête, ont transmis un courrier de soutien.

Les fichiers étaient réceptionnés directement par le Céreq qui a systématiquement vérifié la lisibilité et la conformité des données (année de référence correcte, disponibilité de l'ensemble des informations attendues).

Le Céreq a également :

- organisé le retour d'information auprès du prestataire sur les fichiers réceptionnés (et donc à ne pas relancer). Cette mise à jour a été quotidienne via une base de suivi Céreq ;
- traité les demandes de nouveaux fichiers en cas d'erreur sur le fichier réceptionné (sans repasser par le prestataire) ;
- assuré la codification des diplômes.

Un traitement automatique (via une macro SAS développée en interne) a permis en temps réel de vérifier les variables présentes dans les fichiers importés et les variables obligatoires. Cette macro a notamment permis de réaliser un deuxième travail de correction sur les fichiers déjà visés par l'équipe de réception des fichiers.

Un peu plus de 3 000 établissements ont transmis leurs fichiers, toujours directement au Céreq, soit un taux de réponse de 73 %.

## 2. La compilation et l’apurement de la base de sondage

### a) Les fichiers reçus

**Tableau 3 - Détail de la réception des fichiers par type d'établissement  
(hors bases collectées via les rectorats)**

	Génération 2007 (sans DOM)			Génération 2010 (avec DOM)		
	Nombre total d'établissements	Nombre de fichiers d'établissements reçus	%	Nombre total d'établissements	Nombre de fichiers d'établissements reçus	%
<b>Centres de formation des apprentis</b>	988	777	79	1 064	894	84
<b>Etablissements d'enseignement secondaire agricole</b>	866	670	77	887	746	84
<b>Ecoles de formation de la santé</b>	585	336	57	657	492	75
<b>Ecoles d'architecture et artistiques</b>	332	158	48	350	175	50
<b>Ecoles de commerce et de gestion</b>	280	125	45	312	137	44
<b>Ecoles d'ingénieurs</b>	201	126	63	225	145	64
<b>Ecoles dans le secteur du service</b>	132	47	36	192	76	40
<b>Ecoles de formation sociale</b>	125	75	60	171	126	74
<b>Universités</b>	78	78	100	78	78	100
<b>Dgafp – supplément</b>	26	10	38	40	30	75
<b>Centres de formation pédagogique privés</b>	32	18	56	36	25	69
<b>Ecoles dans le secteur industriel</b>	22	12	55	27	14	52
<b>Ecoles de formation agricole</b>	15	14	93	15	11	73
<b>Centres ou facultés privés</b>	12	6	50	12	7	58
<b>Instituts d'études politiques</b>	11	8	73	10	9	90
<b>Ecoles d'administration publique</b>	6	5	83	6	6	100
<b>Ecoles de la DGA</b>	3	1	33	3	1	33
<b>Ecoles normales supérieures</b>	5	4	80	3	2	67
<b>Instituts universitaires de formation des maîtres</b>	26	19	73			
<b>Total</b>	<b>3 745</b>	<b>2 489</b>	<b>66</b>	<b>4 088</b>	<b>2 974</b>	<b>73</b>

- DRJS : 16 fichiers réceptionnés sur 27 (France + dom) soit 60 %
- Universités : 100 %
- Base Elèves Académique (BEA) : 100 %

Une fois les fichiers uniformisés et vérifiés, l'ensemble des fichiers a été compilé pour constituer une base globale. Une fois constituée, cette base a été apurée pour supprimer, quand cela était possible, les individus hors champ (détectés comme poursuivants en 2010-2011) et les doublons.

## b) Les couplages et apurements des fichiers

Selon les établissements, le Céreq a collecté des individus inscrits, sortants ou diplômés.

**Tableau 4 - Nature des fichiers collectés**

	Collectés auprès des établissements, bases nominatives		Collectés auprès de la Depp et de la Dgesip, bases individuelles mais non nominatives	
	2009-2010	2010-2011	2009-2010	2010-2011
<b>Rectorats</b>	inscrits	inscrits		
<b>Universités</b>	sortants		Sise inscrits et sise couplage : Sortants et inscrits	Sise inscrits
<b>CFA</b>	sortants	inscrits	Sifa inscrits	Sifa inscrits
<b>DRJS - Sports</b>	diplômés sortants			
<b>Ecoles fonction publique + ENS + DGA</b>	diplômés sortants fonctionnaires sortants			
<b>Santé social</b>	diplômés sortants			
<b>Ecoles de commerce</b>	diplômés sortants			
<b>Ecoles d'ingénieurs</b>	diplômés sortants			
<b>Ecoles de formation agricole</b>	diplômés sortants			
<b>IEP</b>	diplômés sortants			
<b>Ecoles dans les secteurs services /industriel</b>	diplômés sortants			
<b>Lycées agricoles, culture, centres et facs privées + centres de formation pédagogiques privés</b>	sortants			

Pour les bases transmises par les rectorats, pour les bases Sifa, et pour les fichiers SISE des inscrits en 2009-2010 et en 2010-2011, la disponibilité d'un identifiant commun de gestion (l'Identifiant National Etudiant ou INE) a permis de repérer de manière assez précise les poursuites d'études internes à ces trois champs, en comparant les fichiers 2009-2010 et les fichiers disponibles pour 2010-2011.

Les poursuites d'études, comme le passage du champ des rectorats ou des centres de formation des apprentis vers l'université ne sont pas toujours repérables à partir de l'INE (car certains INE sont manquants ou provisoires). Une partie des poursuites d'études a toutefois été repérée à partir des



noms, prénoms, dates de naissance (mois, année) entre les fichiers des inscrits en 2009-2010 et les fichiers des inscrits en 2010-2011.

Il reste malgré ces étapes de nombreux individus hors champ dans la base de sondage finale. Les individus poursuivants qui n'ont pas pu être repérés lors de cette étape sont mis hors champ lors de la passation du questionnaire filtre. Pour l'enquête 2013 auprès de la « Génération 2010 », le repérage de personnes hors champ lors de la collecte a concerné 40 849 personnes, soit 52 % des répondants.

Outre les poursuites d'études, des doublons parmi les inscrits de 2009-2010 sont également repérés. Ils le sont à partir des noms, prénoms et dates de naissance (mois, année). Le dé-doublonnage est effectué sur la base complète, tous types d'établissements confondus.

Les individus âgés de plus de 35 ans en 2010 ont également été supprimés, sauf pour les extensions de champ prévues pour les formations du sport et de l'animation et du domaine santé/social ainsi que pour les écoles de la fonction publique de manière à conserver les individus post-initiaux.

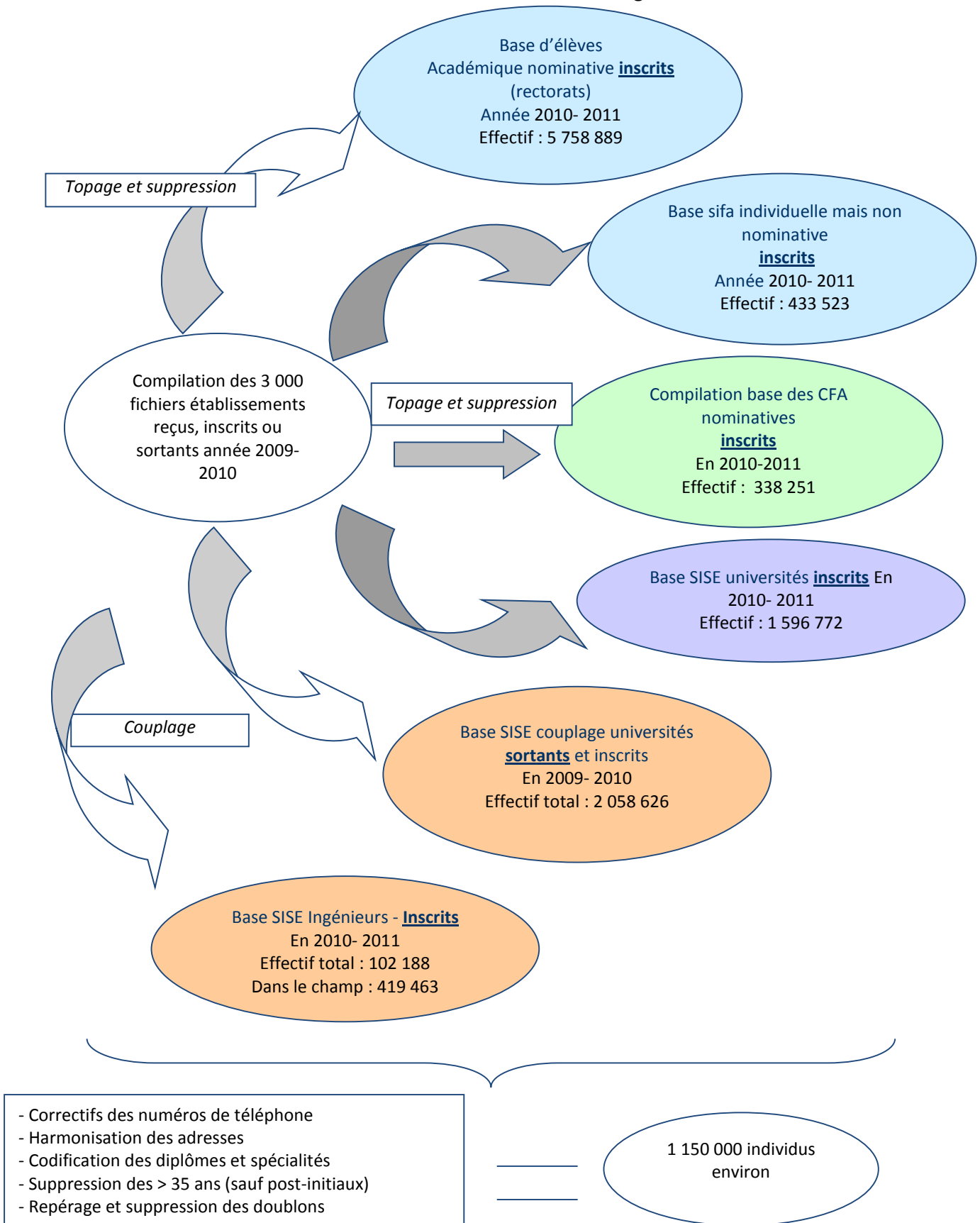
Comme pour « Génération 2007 », les étudiants étrangers ont été conservés. Au final, la base de sondage de l'enquête 2013 auprès de la « Génération 2010 » comportait **1 147 640 individus**. C'est 100 000 individus de moins que pour la précédente Génération, qui s'explique notamment par une suppression plus importante des poursuivants en CFA grâce aux bases SIFA.

Tableau 5 - Effectifs dans la base de sondage par type d'établissement

Type d'établissement	2004	2007	2010	2010 après prise en compte de SIFA*	en %
A - Ecoles d'ingénieurs	14 880	13 842	17 144	17 113	1 %
B - Ecoles de commerce	17 077	13 540	20 183	20 187	2 %
C - Universités	324 959	347 300	333 210	319 617	28 %
D - Ecoles professions sociales	6 345	12 765	7 572	7 574	1 %
E - Ecoles professions de la santé	17 395	22 819	29 984	29 987	3 %
F - DRJS	12 900	13 238	10 547	11 683	1 %
H - IUFM	19 367	21 218	<i>(non collectés)</i>		
J - Ecoles secteur service	1 776	3 161	3 588	3 377	0 %
K - Ecoles secteur industriel	448	2 111	274	251	0 %
L - Ecoles formations agricoles	823	1 093	397	373	0 %
M - Facultés privées	3 793	3 516	4 105	4 002	0 %
N - Lycées et collèges MEN	674 405	604 844	604 287	511 219	45 %
P - Centre de Formation des Apprentis	123 957	149 795	153 803	146 429	13 %
Q - Lycées agricoles	35 557	53 063	58 717	55 862	5 %
R - Centres privés d'enseignement	1 221	947	948	948	0 %
S - Ecoles administrations publiques	250	609	671	671	0 %
T - Ecoles Normales Supérieures	386	816	875	875	0 %
U - Ecoles de la DGA	237	144	167	167	0 %
V - IEP	1 205	2 255	2 446	2 446	0 %
W - Ecoles ministère de la Culture	8 148	9 918	9 530	9 534	1 %
X - CIFRE	430	965	1 133	1 130	0 %
Y - ALLOCATAIRES THESES	2 973	2 712	<i>(désormais comptés dans les universités)</i>		
Z - DGAFP	0	2 682	4 192	4 195	0 %
<b>Total</b>	<b>1 268 532</b>	<b>1 283 353</b>	<b>1 263 774</b>	<b>1 147 640</b>	<b>100 %</b>

\*Chiffres définitifs après suppression des poursuivants en CFA en 2010-2011, issus de la base SIFA n+1

**Figure 4 Les appariements entre bases de données / Estimation du taux de couverture des individus de la base de sondage**



L'estimation de la couverture de la base de sondage, qui renvoie au degré d'exhaustivité de la collecte des fichiers d'établissement, a été effectuée en s'appuyant sur diverses sources administratives. Sauf mention contraire, les taux de couverture présentés portent sur les effectifs couverts et non sur le nombre d'établissements couverts.

*i. Pour le champ « Universités » :*

Les effectifs de la base considérés comme sortants après le repérage des doublons sur le champ université ont été comparés à ceux fournis par le système d'information de suivi des étudiants de la DEPP (système SISE).

Au niveau national, le taux de couverture du champ universitaire est estimé à 86 %.

**Tableau 6 - Estimation de la couverture des universités par niveau (hors IUFM)**

Sortants de...	Effectif collecté par le Céreq	Effectif présent dans SISE	Taux de couverture estimé 2010	Taux de couverture estimé 2007
1 <sup>re</sup> année licence - IUT	7 589	8 587	88 %	88 %
1 <sup>re</sup> année licence universitaire	99 350	111 939	89 %	92 %
2 <sup>de</sup> année licence - IUT	18 679	19 463	96 %	94 %
2 <sup>de</sup> année licence universitaire	25 119	30 058	84 %	90 %
3 <sup>e</sup> année licence professionnelle	29 098	31 017	94 %	88 %
3 <sup>e</sup> année licence universitaire	38 506	43 888	88 %	90 %
1 <sup>re</sup> année de Master	37 240	47 359	79 %	89 %
2 <sup>de</sup> année de Master	70 558	80 673	87 %	91 %
Formations diverses du supérieur	10 279	16 505	62 %	80 %
Doctorats	10 203	13 193	77 %	83 %
Ecoles d'ingénieurs universitaires	4 983	6 423	78 %	73 %
<b>Total</b>	<b>351 604</b>	<b>409 105</b>	<b>86 %</b>	<b>90 %</b>

**Tableau 7 - Estimation de la couverture par discipline**

Discipline	Effectif collecté par le Céreq	Effectif présent dans SISE	Taux de couverture estimé (2010)
Secteur secondaire	31 600	33 600	94 %
Secteur tertiaire	23 800	25 500	93 %
Ecoles d'ingénieurs universitaires	5 000	6 400	78 %
Formations de la santé	10 300	16 500	62 %
Droit, sciences politiques	42 600	50 300	85 %
Economie, gestion	53 700	63 300	85 %
Lettres et arts	53 200	63 400	84 %
Sciences humaines et sociales	53 700	61 600	87 %
Médecine	17 300	19 500	89 %
Sciences fondamentales appliquées	60 400	69 000	88 %
<b>Total</b>	<b>351 600</b>	<b>409 100</b>	<b>86 %</b>

ii. Pour le champ des Centres de Formation par Apprentissage (CFA) :

En termes d'effectifs, le coefficient de couverture dans la base de sondage est de 67 % pour le champ des apprentis. Il a été estimé à partir de l'enquête SIFA<sup>6</sup> sur les centres de formation d'apprentis, enquête annuelle pilotée par la DEPP du ministère de l'Éducation nationale sur l'ensemble des apprentis (CFA)<sup>7</sup>.

**Tableau 8 - Estimation de la couverture du champ CFA**

	Région	Effectif collecté par le Céreq	Effectif présent dans SIFA	Taux de couverture estimé 2010
11	Ile de France	28 108	39 008	72 %
21	Champagne-Ardenne	3 724	4 030	92 %
22	Picardie	4 487	7 589	59 %
23	Haute-Normandie	6 135	6 924	89 %
24	Centre	8 183	9 532	86 %
25	Basse-Normandie	3 458	4 997	69 %
26	Bourgogne	4 750	6 274	76 %
31	Nord-Pas-de-Calais	6 839	9 386	73 %
41	Lorraine	6 360	9 179	69 %
42	Alsace	6 113	6 570	93 %
43	Franche-Comté	3 016	4 468	68 %
52	Pays de la Loire	10 800	15 931	68 %
53	Bretagne	8 348	10 754	78 %
54	Poitou-Charentes	6 065	5 859	100 %*
72	Aquitaine	6 860	9 407	73 %
73	Midi-Pyrénées	6 522	10 759	61 %
74	Limousin	1 868	2 270	82 %
82	Rhône-Alpes	18 692	23 550	79 %
83	Auvergne	4 361	4 266	100 %
91	Languedoc-Roussillon	4 402	7 489	59 %
93	Paca	12 466	20 353	61 %
94	Corse	775	840	92 %
97	Dom	2 257	3 303	68 %
	<b>TOTAL</b>	<b>164 589</b>	<b>222 738</b>	<b>74 %</b>

\*il arrive parfois que les effectifs des fichiers collectés soit supérieurs aux effectifs des sources de référence. On considère alors que la couverture est de 100 %.

<sup>6</sup> SIFA : Système d'Information sur la Formation des Apprentis

<sup>7</sup> L'enquête fournit, sur différents échelons géographiques, des statistiques sur les effectifs par sexe, âge, niveau et spécialité de formation, secteur d'activité de l'employeur (code APE). Les données sont collectées de façon agrégée.

iii. Pour le champ des formations « Jeunesse et Sport » :

Le ministère des Sports a constitué une base nationale anonyme à partir d'informations issues des directions régionales de la jeunesse et des sports. Cette base existe depuis peu, elle est destinée à devenir nominative. Le taux de couverture estimé est de 69 % en 2010.

**Tableau 9 - Estimation de la couverture du champ sport**

Diplôme	Nombre d'individus dans la base de sondage	Nombre d'individus dans la base nationale du ministère	Taux de couverture 2010
BAPAAT (Brevet d'aptitude professionnelle d'assistant animateur technicien de la jeunesse et des sports)	513	836	61 %
BEES (Brevet d'état d'éducateur sportif)	3 426	4 693	73 %
BPJEPS (Brevet professionnel de la jeunesse, de l'éducation populaire et du sport)	5 470	8 690	63 %
DEFA (Diplôme d'état relatif aux fonctions d'animation)	28	86	33 %
DESJEPS, DEJEPS (Diplôme d'état –supérieur- de la jeunesse, de l'éducation populaire et du sport)	3 040	3 648	83 %
<b>Total</b>	<b>12 477</b>	<b>17 953</b>	<b>69 %</b>

iv. Pour le champ des formations du domaine de la Santé et du Social :

Le champ a été défini à partir d'une liste d'établissements transmise par la DREES. Au total, on comptabilisait 37 561 diplômés des écoles de la santé et du social en 2010 dans la base de sondage (25 378 pour 2007).

Selon le fichier national des établissements sanitaires et sociaux (FINESS)<sup>8</sup>, ils étaient 61 252, soit une couverture moyenne de 61 % pour la base de sondage en termes d'effectifs (contre 40 % pour Génération 2007). Le taux de couverture a fortement augmenté, grâce à une collecte plus intensive sur ce champ.

<sup>8</sup> Le fichier FINESS est un fichier géré par le ministère en charge de la santé et des affaires sociales pour répondre d'abord à des besoins de gestion, puisqu'il permet l'immatriculation par les DDASS et les DRASS des établissements et entités juridiques porteurs d'une autorisation ou d'un agrément dans les trois domaines : le sanitaire et médico-social, le social, l'enseignement des personnels en santé et en social.

**Tableau 10 - Estimation de la couverture du champ Santé social**

Détail		Effectif collecté	Effectif total	Taux de couverture moyen estimé
414	Infirmier	14 871	21 641	69 %
416	Masseur-kiné	1 245	2 015	62 %
420	Pédicure podologue	87	449	19 %
421	Manipulateur radio	222	591	38 %
423	Ergothérapeute	228	371	61 %
424	Psychomotricien	103	515	20 %
415	Sage-femme	256	902	28 %
418	Puéricultrice	314	1 007	31 %
419	Aide-soignant	10 853	17 974	60 %
456	Auxiliaire puériculture	1 484	3 447	43 %
422	Conseiller économique, social et familial	700	1 229	57 %
432	Assistant service social	1 446	2 256	64 %
435	Educateur jeunes enfants	935	1 214	77 %
433	Educateur spécialisé	2 686	4 244	63 %
437	Moniteur éducateur	1 723	2 487	69 %
	DOM	408	910	45 %
	<b>Total</b>	<b>37 561</b>	<b>61 252</b>	<b>61 %</b>

v. *Pour le champ des Instituts Universitaires de Formation des Maîtres (IUFM) :*

La réforme de mastérisation des concours d'enseignement, mise en place sous régime transitoire l'année 2010, et surtout le rattachement administratif des IUFM aux universités, n'ont pas permis d'intégrer dans cette enquête Génération les sortants de 1<sup>re</sup> année de cette formation.

vi. *Pour le champ des écoles de commerce et d'ingénieurs :*

La référence pour le calcul du taux de couverture est issue de la publication "Repères et références statistiques 2012" (RERS 2012) de la Depp. Les chiffres portent sur le nombre de diplômés d'écoles de commerce et ingénieurs en France métropolitaine + DOM en 2010.

**Tableau 11 - Estimation de la couverture des écoles d'ingénieurs et des écoles de commerce**

Détail	Effectif collecté	Effectif total	Taux de couverture estimé (2010)	Taux de couverture estimé (2007)
Ecoles d'Ingénieurs	17 144	28 928	59 %	49 %
Ecole de commerce	20 183	28 852	70 %	56 %

vii. *Pour les établissements sans information auxiliaire :*

Pour l'ensemble des établissements suivants, nous n'avons trouvé aucune référence statistique fiable pour calculer la couverture. Un total de 82 700 individus (dont les lycées agricoles) présents dans la

base de sondage sont concernés. Le taux de couverture est dans ce cas obtenu en appliquant le taux de réponse de la collecte établissement classé par type.

**Tableau 12 - Taux de couverture établissement des autres établissements**

Type d'établissement	Libellé	Nombre d'établissements répondants	Nombre d'établissements sollicités	Taux de couverture estimé
J	Ecoles dans le secteur services	76	192	40 %
K	Ecoles dans le secteur industriel	14	27	52 %
L	Ecoles de formation agricole	11	15	73 %
M	Centres ou facultés privés	7	12	58 %
Q	Lycées agricoles	641	887	72 %
R	Centres de formation pédagogiques privés	25	36	69 %
S	Ecoles d'administration publique	6	6	100 %
T	Ecoles Normales Supérieures	2	3	67 %
U	Ecoles de la DGA	1	3	33 %
V	Instituts d'études politiques	9	10	90 %
W	Ecoles d'architecture et artistiques	175	350	50 %
Z	Ecoles collectées pour la DGAFP	30	40	75 %
<b>TOTAL</b>		<b>997</b>	<b>1 581</b>	



## viii. Récapitulatif

Tableau 13 - Tableau récapitulatif des taux de couverture

Type d'établissement	Nombre d'individus dans la base de sondage	Taux de couverture moyen estimé
A - Ecoles ingénieurs	17 110	0,59
B - Ecoles de commerce	20 190	0,70
C - Universités	319 620	0,86
D - Ecoles professions sociales	7 570	0,65
E - Ecoles professions de la santé	29 990	0,63
F - DRJS	11 680	0,69
J - Ecoles secteur service	3 380	0,40
K - Ecoles secteur industriel	250	0,52
L - Ecoles formations agricoles	370	0,73
M - Facultés privées	4 000	0,58
N - Lycées et collèges MEN	511 220	1
P - Centre de Formation des Apprentis	146 430	0,74
Q - Lycées agricoles	55 860	0,72
R - Centres privés d'enseignement	950	0,69
S - Ecoles administrations publiques	670	1
T - Ecoles Normales Supérieures	870	0,67
U - Ecoles de la DGA	170	0,33
V - IEP	2 450	0,9
W - Ecoles ministère de la Culture	9 530	0,5
X - CIFRE	1 130	1
Z - DGAFP	4 200	0,75
<b>TOTAL</b>	<b>1 147 640</b>	<b>0,85</b>

## c) Amélioration de la qualité de la base de sondage : les numéros de téléphones, les mails

Pour la base de sondage de l'enquête 2013 auprès de la « Génération 2010 », l'effort pour obtenir davantage de numéros de téléphone a été poursuivi.

Le nombre de champs concernant les numéros de téléphone a été augmenté à partir de la Génération 2007 : jusqu'à 4 numéros par individu. Les numéros de téléphone portable ont été mentionnés dans les structures de fichier pour insister sur cette information.

Au total, 82 % des individus présents dans la base de sondage ont au moins un numéro de téléphone, contre 75 % pour la Génération 2007. Le nombre moyen de numéros de téléphone dans la base de sondage pour les individus ayant au moins un numéro est estimé à 1,3 pour l'enquête 2013 auprès de la « Génération 2010 ». Cet indicateur a peu évolué.

**Tableau 14 - Evolution du nombre de numéros de téléphone selon l'enquête**

Génération	Individus ayant au moins un numéro dans la base de sondage
2010	82 %
2007	75 %
2004	27 %

**Tableau 15 - Fréquence de numéros de téléphone selon le type**

Numéro de téléphone Fixe	Numéro de téléphone portable (06)	Numéro de téléphone spéciaux (07,08,09)	Nombre d'occurrences	Pourcentage
0	0	1	13 238	1 %
0	0	2	275	0 %
0	1	0	198 338	21 %
0	1	1	8 256	1 %
0	1	2	19	0 %
0	2	0	15 447	2 %
0	2	1	228	0 %
0	2	2	2	0 %
0	3	0	294	0 %
0	4	0	1	0 %
1	0	0	412 050	44 %
1	0	1	2 655	0 %
1	0	2	2	0 %
1	1	0	250 543	27 %
1	1	1	358	0 %
1	2	0	6 783	1 %
1	2	1	10	0 %
1	3	0	15	0 %
2	0	0	26 727	3 %
2	0	1	9	0 %
2	1	0	3 332	0 %
2	1	1	1	0 %
2	2	0	87	0 %
3	0	0	113	0 %
3	1	0	7	0 %
4	0	0	1	0 %
			<b>938 791</b>	<b>100 %</b>

*Note de lecture : Dans la base de sondage, 27 % des individus ayant au moins un numéro de téléphone ont un numéro de fixe normal et un numéro de portable (contre 17,2 % pour la génération 2007).*

Les jeunes de la Génération 2010 disposent en moyenne d'1,3 numéro de téléphone par individu.

Pour les individus pour lesquels la base ne dispose d'aucun numéro de téléphone (mais également pour ceux qui en disposent), une recherche téléphonique est réalisée par un prestataire externe. La récupération de plusieurs numéros est autorisée en cas de doute, l'identification étant faite lors de la collecte.

#### d) Le géocodage de la base de sondage

L'expérience de l'enquête 2010 auprès de la "Génération 2007", a mis en évidence que les jeunes résidant en Zones Urbaines Sensibles (ZUS) à la fin de leurs études tendaient à moins répondre à l'enquête que les autres jeunes. Population aux conditions d'insertion spécifiques, il apparaît donc nécessaire d'intégrer cette dimension dès la phase d'échantillonnage des enquêtes Génération. La base de sondage de l'enquête 2013 auprès de la « Génération 2010 » a donc été géocodée. Le géocodage a été réalisé à partir des adresses des parents/jeunes à la sortie de leur formation en 2009-2010.

Ainsi, nous identifions 6,8 % des jeunes de la base de sondage qui résidaient en ZUS l'année de la sortie du système éducatif (ou dont les parents résidaient en ZUS).

Cette variable est utilisée dans le calcul des probabilités de tirage dans l'échantillon.

### III. Le plan de sondage et la constitution de l'échantillon

#### 1. Objectifs et difficultés du plan de sondage

Le plan de sondage de l'enquête 2013 auprès de la « Génération 2010 » vise à l'obtention de 40 000 questionnaires, 25 000 questionnaires pour les besoins du Céreq et environ 15 000 pour les différentes demandes d'extension.

L'échantillon doit satisfaire les contraintes suivantes :

- Permettre de réaliser un nombre minimum de questionnaires équilibrés par rapport aux formations de sortie
- Surreprésenter certains types de formations sur lesquelles le Céreq souhaite réaliser une analyse fine ou sur lesquels l'attrition lors des ré-interrogations est particulièrement élevée
- Permettre de satisfaire les besoins des partenaires d'extensions
- Permettre de disposer d'une réserve, utilisable de manière totale ou partielle si les taux de réponse observés en cours d'enquête se révèlent inférieurs aux prévisions.

#### 2. Description du plan de sondage

Le tirage est stratifié par région de formation et équilibré par type et spécialité de formation. Il s'agit d'un tirage à probabilités inégales.

Le tirage est réalisé à l'aide de l'algorithme du « Cube » (Deville-Tillé, 2004) et de la macro-procédure Sas associée, développée par l'Insee. Cet algorithme d'échantillonnage permet d'améliorer la précision des estimateurs associés aux variables d'intérêt de l'enquête. Pour cela, Cube procède au tirage aléatoire d'un échantillon équilibré sur un ensemble de totaux connus à partir d'informations auxiliaires. La méthode consiste à choisir un échantillon tel que les estimateurs d'Horvitz-Thompson des totaux des variables choisies pour l'équilibrage, coïncident avec les vrais totaux. Cette procédure s'applique sur des bases de sondage munies d'informations auxiliaires, qualitatives ou quantitatives, connues au niveau individuel. Des probabilités d'inclusion individuelles destinées à surreprésenter les individus issus de certains types de formation ou inclus dans des champs d'extensions sont intégrées dans la procédure de tirage. A titre indicatif, l'ordre de grandeur de l'échantillon, réserve comprise, devrait se situer aux alentours de 300 000 individus.

##### a) Travail préparatoire

Dans un premier temps, il s'agit de déterminer les probabilités individuelles permettant la réalisation de 20 000 questionnaires représentatifs de l'ensemble des sortants du système éducatif en 2010. Dans un second temps, des coefficients de supplément de tirage permettront de surreprésenter des populations particulières.

**Première étape :** Pour obtenir une structure des répondants qui soit proche de la structure des sortants dans la population générale, le calcul des probabilités d'inclusion des individus de la base de sondage a pris en compte les probabilités de répondre anticipées, modélisées par les caractéristiques observables des individus sur les précédentes enquêtes Génération, et les taux de couverture relatifs aux établissements dont les individus sont issus.

En effet, la prise en compte dans le calcul de la probabilité d'inclusion d'un coefficient lié à la probabilité de répondre anticipée permet d'éviter que les individus qui répondent le moins soient sous représentés dans l'ensemble des répondants. De même, pour éviter que les sortants issus d'établissements mal couverts par notre base de sondage soient sous représentés par rapport à la

structure générale de la population, la probabilité d'inclusion dans l'échantillon prend aussi en compte un coefficient lié au taux de couverture. Ces deux coefficients ne sont pas suffisants pour déterminer la probabilité de tirage final. Pour approcher le nombre de questionnaires voulu (20 000), il faut enfin déterminer un coefficient multiplicateur global (ou coefficient de dilatation).

**Seconde étape :** En fonction de la différence entre le nombre de questionnaires souhaité sur chaque sous-population d'intérêt et le nombre attendu avec les probabilités de tirage calculées précédemment, des coefficients multiplicatifs sont appliqués à chaque individu appartenant à une sous population d'intérêt.

*i. La création d'une variable de pilotage pour le plan de tirage*

L'échantillonnage suppose d'estimer pour chaque individu présent dans la base de sondage une probabilité de réponse, en fonction de ses caractéristiques propres. On sait par avance que la propension à répondre est liée au type de formation suivie (niveau, spécialité, type d'établissement, etc.). La variable strate disponible dans la base de sondage fournit une indication détaillée sur la formation de sortie. Elle contient environ 200 modalités et les effectifs par modalité sont très disparates. Bien que très fine, la variable strate ne tient pas compte de la spécialité de formation, ce qui peut constituer un inconvénient. Il convient de la retravailler (en agrégeant les modalités trop peu représentées, en tenant compte de la spécialité au besoin) afin qu'elle puisse être utilisée comme variable explicative lors de la modélisation de la non réponse. Cette variable est également construite au regard du plan de sondage (nombre total de questionnaires attendus, extensions, etc.).

Etant donné que le type de formation suivie est également lié aux variables caractérisant l'insertion (taux d'emploi, type d'emploi occupé, niveau de rémunération, etc.), elle peut également être utilisée comme variable d'équilibrage afin d'améliorer les propriétés du tirage.

Pour « Génération 2004 » et « Génération 2007 », les variables de pilotage ont été respectivement *strateg* (128 modalités) et *strate07* (27 modalités). Pour « Génération 2010 », la variable de pilotage se nomme *stradef* et comporte 53 modalités.

Au départ, on travaille sur la variable *stra00*, construite à partir du croisement de la variable *strate* avec la spécialité de formation. Lorsque les effectifs sont trop faibles, il n'y a pas de distinction par spécialité.

Pour chaque modalité de *stra00*, on calcule les résultats suivants obtenus à partir de « Génération 2004 » :

- taux de réponse au questionnaire filtre ;
- taux de réponse dans le champ ;
- ventilation des répondants par niveau de plus haut diplôme ;
- taux de chômage à 3 ans ;
- taux d'emploi à 3 ans ;
- taux d'EDI à 3 ans ;
- taux de cadres-professions intermédiaires à 3 ans ;
- part d'hommes ;
- taux de réponse entre 3 et 5 ans ;
- taux de réponse entre 5 et 7 ans.

On effectue ensuite une classification ascendante hiérarchique (CAH), en s'appuyant d'une part sur les classes proposées par la CAH et en restant cohérents d'autre part avec les niveaux de sortie et

structures de formation. À partir de la classification obtenue, on construit la variable stradef de manière raisonnée.

*ii. La modélisation de la probabilité de compléter un questionnaire dans le champ*

Pour chaque individu présent dans la base de sondage, il s'agit d'estimer :

- la probabilité de répondre au questionnaire filtre ;
- la probabilité d'appartenir au champ de l'enquête.

On va s'appuyer sur ce qui a été observé sur les enquêtes précédentes, à savoir « Génération 2004 » et « Génération 2007 ». Pour estimer la probabilité d'appartenir au champ de l'enquête, on tiendra également compte des améliorations réalisées lors de la constitution de la base de sondage de « Génération 2010 » (meilleure détection des individus hors champ).

La variable stradef est reconstituée a posteriori sur les échantillons de « Génération 2004 » et « Génération 2007 ».

*iii. Le taux de réponse au questionnaire filtre*

Génération 2004

L'échantillon utilisé pour « Génération 2004 » était constitué de 371 500 individus dont 115 800 ont répondu au questionnaire filtre, portant le taux de réponse à 31,2 %.

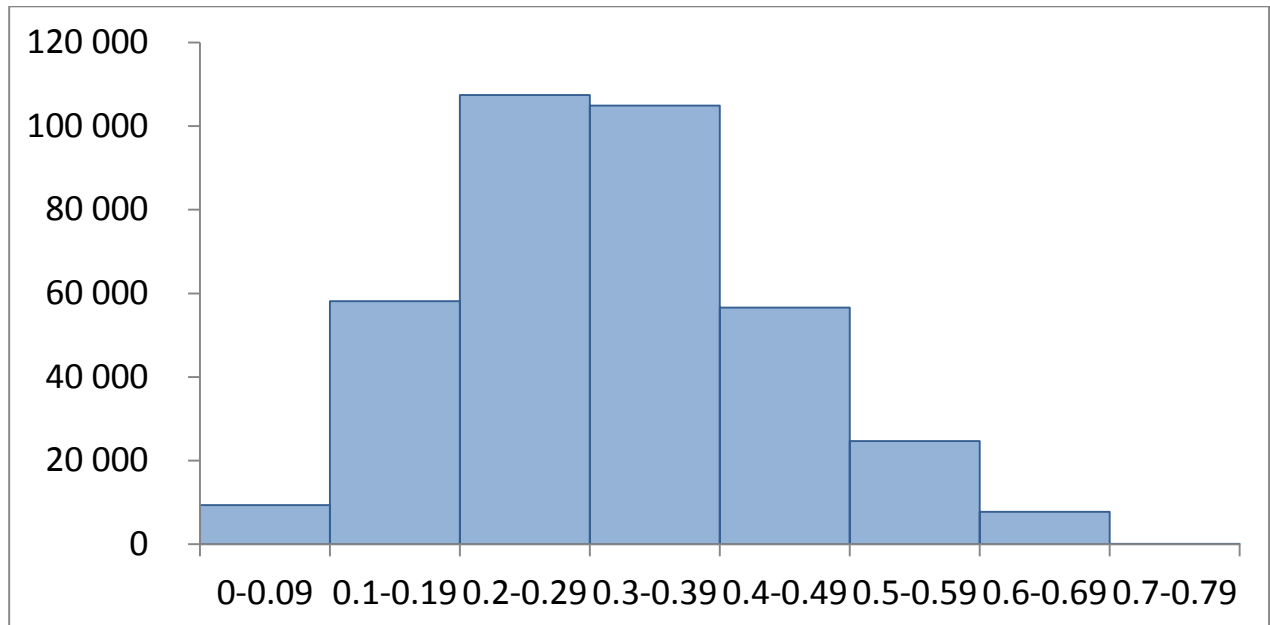
La probabilité de répondre au questionnaire filtre a été estimée par régression logistique, les variables explicatives étant les suivantes :

- région de l'établissement de formation (region) ;
- genre de l'individu (sexe) ;
- strate de formation en 2004 (stradef) ;
- formation effectuée en apprentissage en 2004 (apprent) ;
- type de coordonnées téléphoniques dans la base de sondage ;
- commune de l'adresse postale en 2004 renseignée (icommune) ;
- code postal en 2004 renseigné (icp).

A caractéristiques comparables, les jeunes formés en Ile-de-France, Languedoc-Roussillon et Corse sont ceux qui présentent les taux de réponse les plus faibles, a contrario de ceux formés en Limousin, Bretagne ou Pays de la Loire. Les individus possédant un numéro de téléphone portable dans la base de sondage ont un taux de réponse globalement élevé. La propension à répondre dépend également beaucoup du type de formation suivie, les thésards constituant la population qui répond le moins. Les jeunes résidant dans un pôle urbain répondent moins que les autres.

En moyenne, le taux de réponse estimé sur l'échantillon est de 0,316 (min=0,004 ; max=0,728 ; ET=0,126). On remarque que l'estimation est légèrement supérieure au taux de réponse observé (0,312).

**Figure 5 - Distribution de la probabilité estimée de réponse au questionnaire filtre au sein de l'échantillon de « Génération 2004 »**



#### Génération 2007

L'échantillon utilisé pour « Génération 2007 » était constitué de 191 900 individus dont 48 500 ont répondu au questionnaire filtre, portant le taux de réponse à 25,3 %.

La probabilité de répondre au questionnaire filtre a été estimée par régression logistique, les variables explicatives étant les suivantes :

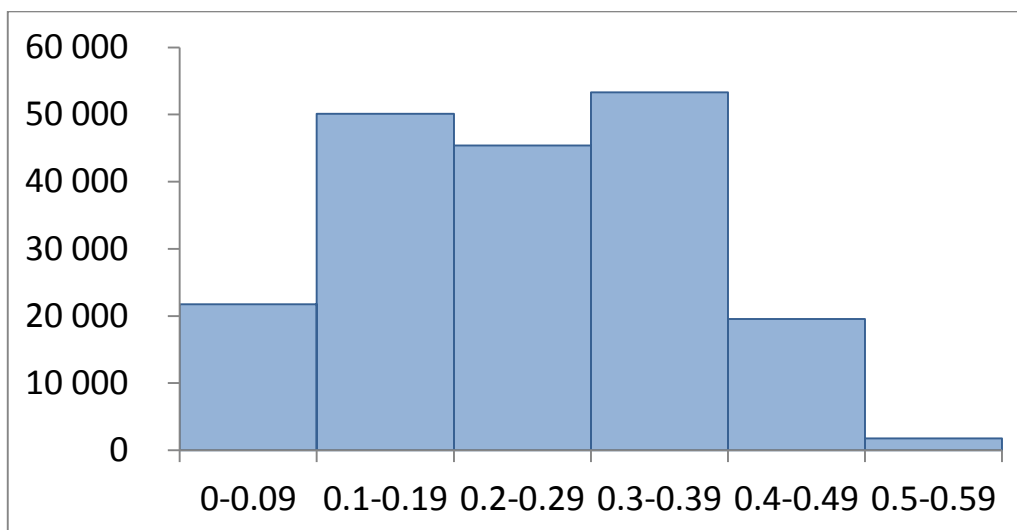
- région de l'établissement de formation (region) ;
- strate de formation en 2007 (stradef) ;
- formation effectuée en apprentissage en 2007 (apprent) ;
- type de coordonnées téléphoniques dans la base de sondage ;
- commune de l'adresse postale en 2007 renseignée (icommune) ;
- code postal en 2007 renseigné (icp) ;
- réside en ZUS en 2007 (ZUS).

Le genre n'a pas été retenu dans le modèle logistique car il n'apparaît pas lié à la probabilité de répondre.

A caractéristiques comparables et comme pour « Génération 2004 », les jeunes formés en Ile-de-France, Languedoc-Roussillon sont ceux qui présentent les taux de réponse les plus faibles, a contrario de ceux formés en Limousin, Bretagne ou Pays de la Loire. Les individus possédant un numéro de téléphone portable et un numéro de téléphone fixe dans la base de sondage ont un taux de réponse globalement élevé. La propension à répondre dépend également beaucoup du type de formation suivie, les thésards constituant à nouveau la population qui répond le moins. Les jeunes résidant en ZUS ou dans un pôle urbain répondent moins que les autres.

En moyenne, le taux de réponse estimé sur l'échantillon est de 0,253 (min=0,007 ; max=0,596 ; ET=0,118).

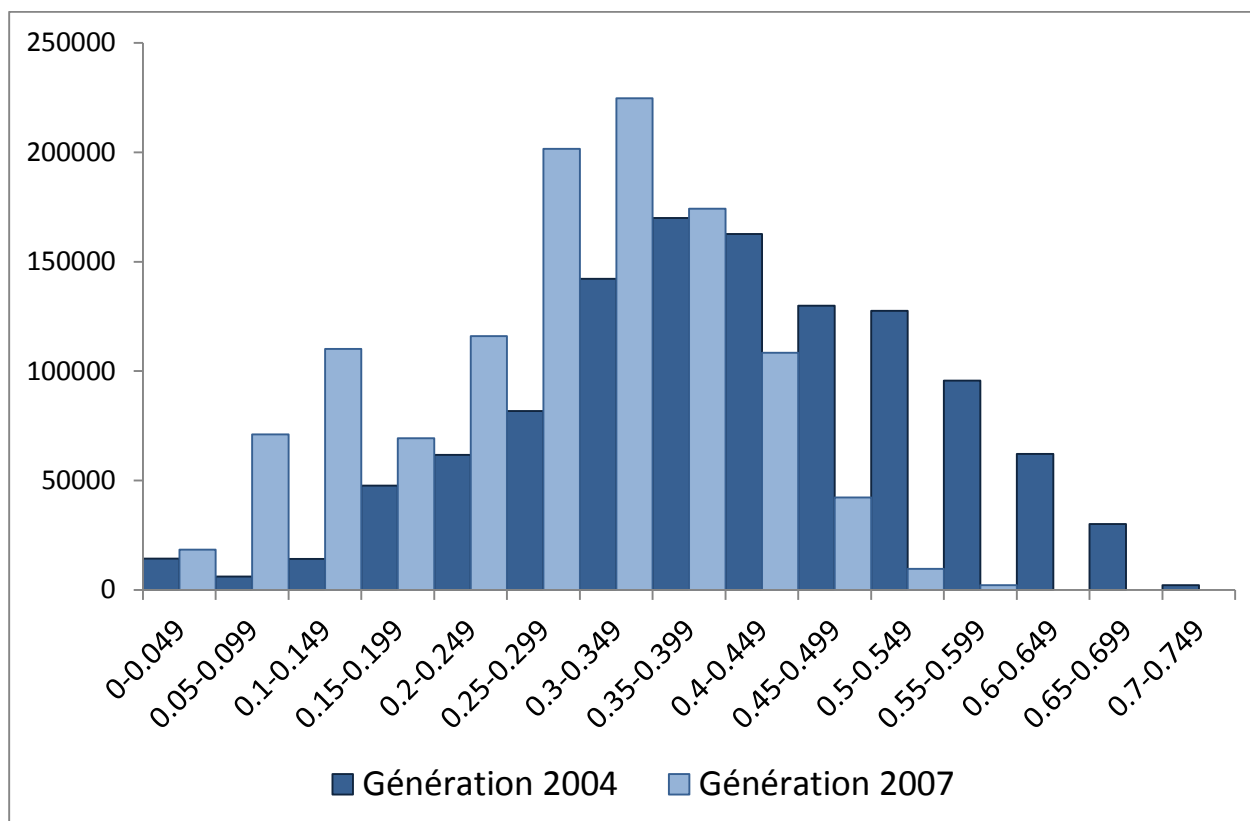
**Figure 6 - Distribution de la probabilité estimée de réponse au questionnaire filtre au sein de l'échantillon de « Génération 2007 »**



*iv. Coefficient retenu*

Les variables retenues pour modéliser la probabilité de réponse des enquêtes « Génération 2004 » et « Génération 2007 » sont présentes dans la base de sondage de « Génération 2010 ». Pour chaque individu de la base de sondage, il est donc possible d'estimer la probabilité de réponse, selon l'hypothèse d'un comportement similaire à « Génération 2004 » ou à « Génération 2007 ».

**Figure 7 - Distribution de la probabilité estimée au sein de la base de sondage de « Génération 2010 », à partir des résultats de « Génération 2004 » et « Génération 2007 »**





La probabilité moyenne de réponse estimée sur la base de sondage est de 0,410 (min=0,002, max=0,739, ET=0,136) sous l'hypothèse d'un comportement de réponse similaire à « Génération 2004 ». Elle est de 0,283 (min=0,007, max=0,600, ET=0,111) sous l'hypothèse d'un comportement de réponse similaire à « Génération 2007 ».

On constate qu'en moyenne, la probabilité de répondre estimée sur la base de sondage est supérieure à celles constatées sur les échantillons de « Génération 2004 » et « Génération 2007 ». Cela traduit le fait que les structures base de sondage/échantillon, diffèrent.

Le choix d'un coefficient estimateur de la probabilité de répondre est stratégique. Nous disposons pour ce faire des éléments suivants.

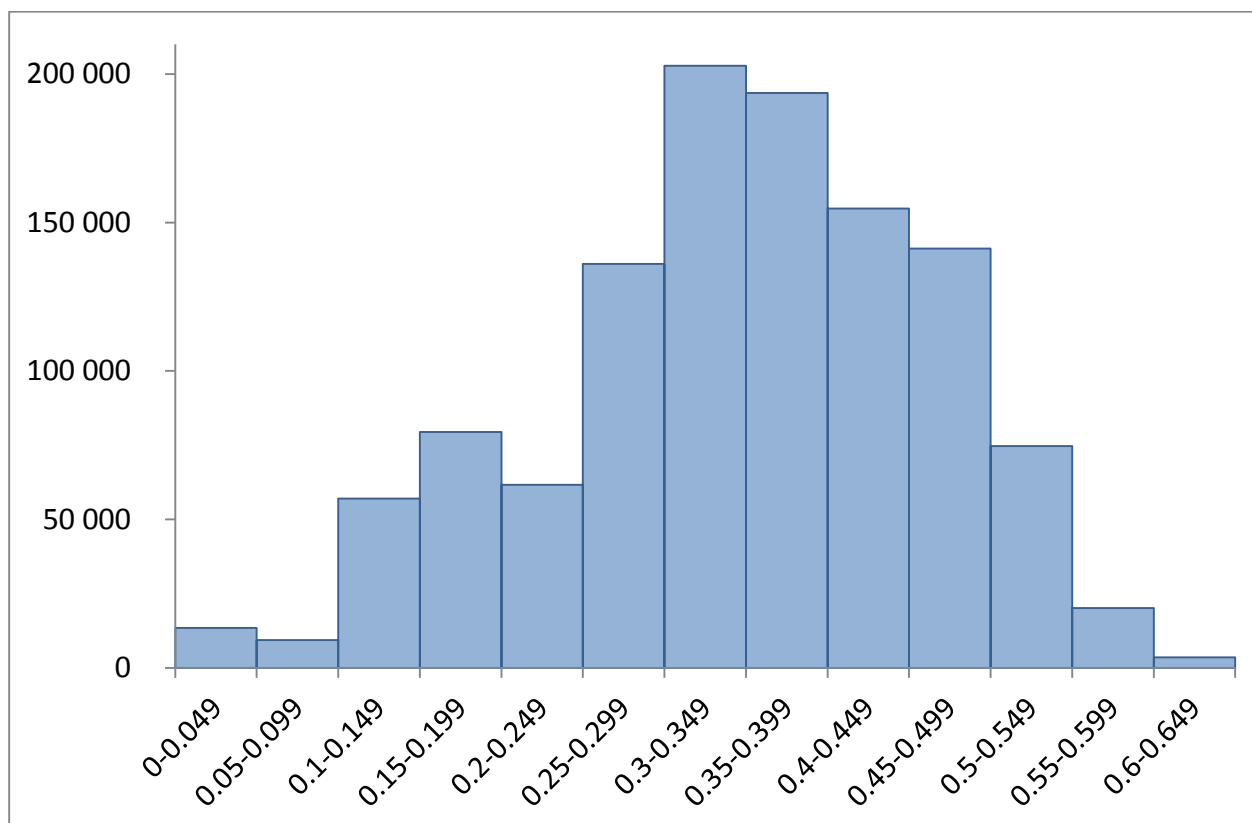
« Génération 2007 » est la plus récente tandis que « Génération 2004 » présente un échantillon plus conséquent et mieux équilibré. Le pouvoir explicatif du modèle de « Génération 2007 » semble légèrement supérieur.

Le taux de réponse moyen a nettement chuté entre les deux générations. D'une part, on sait qu'en règle générale, les taux de réponse ont tendance à diminuer au cours du temps. D'autre part, l'exploitation de l'échantillon de « Génération 2007 » a été moins poussée que celle de « Génération 2004 » (moins de tentatives d'appels téléphoniques).

Finalement, on choisit de retenir comme probabilité estimée la moyenne des estimations de « Génération 2004 » et « Génération 2007 ». Cela permet de lisser les estimations et de rendre l'estimation plus robuste. Notons qu'il s'agit d'une hypothèse volontariste sur le taux de réponse puisque concrètement, on vise un meilleur taux de réponse que pour « Génération 2007 ».

La probabilité moyenne de réponse ainsi estimée sur la base de sondage est de 0,346 (min=0,005, max=0,641, ET=0,117).

**Figure 8 - Distribution de la probabilité estimée retenue au sein de la base de sondage de « Génération 2010 »**



### v. Estimation de la probabilité d'appartenir au champ de l'enquête

Tous les individus répondants ne donnent pas lieu à la réalisation d'un questionnaire. En effet, une partie d'entre eux va se révéler hors du champ de l'enquête lors de la passation du questionnaire filtre. Afin d'estimer le nombre de questionnaires réalisés sur un échantillon, il faut donc pour chaque individu, estimer d'une part, sa probabilité de répondre (au questionnaire filtre) et d'autre part, sa probabilité d'appartenir au champ de l'enquête.

Pour chaque strate de formation (stradef), on a calculé la probabilité d'appartenir au champ des répondants de « Génération 2004 » d'une part, et des répondants de « Génération 2007 », d'autre part. On a ensuite retenu comme première estimation de la probabilité d'appartenir au champ la moyenne pondérée (par le nombre de répondants de la strate à chaque enquête).

Pour « Génération 2010 », un plus grand nombre d'individus a pu être repéré comme ne faisant pas partie du champ de l'enquête (grâce à un appariement de fichiers) lors de la constitution de la base de sondage. On a tenu compte de ce gain dans l'estimation de la probabilité d'appartenir au champ pour chaque strate.

Les collégiens et lycéens généraux sont ceux qui présentent les plus faibles probabilités d'appartenir au champ (moins de 20 %) tandis que les jeunes formés en licences professionnelles, les thésards ou les bacs professionnels de spécialité industrielle sont le plus souvent dans le champ de l'enquête (environ 80 %).

### b) Simulation de 20 000 questionnaires « représentatifs » Céreq.

Cette étape consiste à calculer pour chaque individu présent dans la base de sondage, un coefficient de tirage permettant de simuler un échantillon correspondant à 20 000 individus répondants dans le champ de l'enquête, respectant en proportion la répartition des effectifs de sortants.

Pour calculer le coefficient de tirage, on a mobilisé les taux de couverture estimés et les taux de réponses estimés, en fonction des caractéristiques de l'individu.

Enfin, il s'agit de déterminer le coefficient d'échantillonnage, le même pour l'ensemble des individus de la base de sondage, permettant la réalisation de 20 000 questionnaires dans le champ de l'enquête du Céreq.

Le coefficient de tirage de l'individu  $i$  se décompose de la manière suivante :

$TX1(i) = TC(i) * TNR(i) * TE$ , avec :

- $TC$ , le coefficient lié à la couverture ;
- $TNR$ , le coefficient lié à la non-réponse attendue au questionnaire filtre ;
- $TE$ , le coefficient d'échantillonnage.

### c) Simulation de 25 000 questionnaires Céreq.

La probabilité de tirage précédemment calculée vise à assurer la réalisation de 20 000 questionnaires « représentatifs ».

Il s'agit maintenant d'augmenter le coefficient de tirage de certains individus présentant des caractéristiques spécifiques (formés dans des petites régions, bas niveaux de qualifications, faible taux de réponse aux ré-interrogations...).

Nous avons tout d'abord souhaité surreprésenter les individus susceptibles de présenter un faible taux de réponse lors des ré-interrogations. On a modalisé la probabilité de répondre à 5 ans des répondants dans le champ de « Génération 2004 » grâce à une régression logistique, puis on a attribué cette probabilité aux individus de la base de sondage de « Génération 2010 ». En moyenne, leur probabilité de répondre à 5 ans (sachant qu'ils sont dans le champ de l'enquête et qu'ils ont répondu à 3 ans) est

de 58 %. Pour tous les individus de la base de sondage qui avaient une probabilité estimée de réponse à 5 ans inférieure à 60 %, nous avons établi un coefficient de supplément de tirage tenant compte du delta entre cette probabilité estimée à 60 % et permettant la réalisation de 5 000 questionnaires supplémentaires.

Des suppléments de tirage relatifs aux régions de formation ont ensuite été calculés. Pour chaque région métropolitaine (hormis la Corse pour laquelle cet objectif n'est pas atteignable), l'objectif est d'obtenir un minimum de 250 questionnaires pour les sortants de l'enseignement secondaire et de 250 questionnaires pour les sortants de l'enseignement supérieur.

Les individus formés dans les régions pour lesquelles l'estimation du nombre de questionnaires à partir des probabilités de tirage calculées précédemment est inférieure à ce chiffre, se voient attribuer un coefficient de supplément de tirage :

$$st\_Rk(i) = \frac{250}{Nb \text{ de questionnaires estimé suite à la première étape dans la catégorie } Rk} - 1 \text{ si}$$

l'individu  $i$  appartient à la catégorie  $Rk$

$$st\_Rk(i) = 0 \text{ sinon.}$$

Pour un individu présent dans la base de sondage, le taux de tirage devient :

$$TX2(i) = TX1(i) * (1 + st\_rep5ans(i) + st\_Regniveau(i))$$

A l'issue de cette étape, les coefficients de tirage des individus présents dans la base de sondage ont été calibrés pour viser la réalisation d'environ 30 000 questionnaires dans le champ de l'enquête Céreq.

#### d) Calcul des suppléments de tirage pour les extensions et coefficient de tirage final

Les suppléments de tirage relatifs à chaque extension, par catégorie fine, ont été calculés de la même manière que précédemment :

On détermine un coefficient de supplément de tirage

$$st\_Ek(i) = \frac{Nb \text{ questionnaires souhaités dans la catégorie } Ek}{Nb \text{ de questionnaires estimé suite à la seconde étape dans la catégorie } Ek} - 1 \quad \text{si}$$

l'individu  $i$  appartient à la catégorie  $Ek$

$$st\_Ek(i) = 0 \text{ sinon.}$$

Autant de coefficients de supplément de tirage ont été calculés qu'il y a de catégories d'extensions. Pour un individu présent dans la base de sondage, le poids de tirage devient :

$$TX3a(i) = TX2(i) * (1 + st\_E1(i) + \dots)$$

Pour certains individus, l'application des coefficients de supplément de tirage liés aux extensions conduit à un taux de tirage supérieur à 1 ( $TX3a(i) > 1$ ). Le coefficient de tirage est alors ramené à un et l'individu est automatiquement échantillonné.

Le coefficient de tirage correspondant à l'échantillon principal est  $TX3(i) = \text{Min}[TX3a(i), 1]$ .

### e) Tirage de l'échantillon Global (y compris réserve)

Le tirage de l'échantillon définitif a été réalisé en deux temps. Un premier tirage a permis de sélectionner l'échantillon, réserves comprises. Le second tirage a permis de distinguer, au sein de cet échantillon complet, l'échantillon principal de la réserve.

Le tirage de l'échantillon définitif, réserves comprises, a été effectué en respectant les contraintes de stratification et d'équilibrage.

Les suppléments de tirage utilisés pour permettre la constitution de la réserve sont les suivants :

- $TX4(i) = \min(TX3(i) * 4/3, 1)$  si l'individu appartient à une extension ou s'il a été formé dans un DOM
- $TX4(i) = \min(TX3(i) * 100/85, 1)$ , sinon.

Le tirage a été réalisé sous le logiciel SAS avec l'aide de la macro procédure CUBE<sup>9</sup> développée par l'INSEE.

### f) Tirage de l'échantillon principal

La probabilité conditionnelle d'appartenir à l'échantillon principal sachant qu'on appartient à l'échantillon complet est la suivante :

- $TX3(i)/TX4(i)$  si  $TX4(i) < 1$
- $TX3(i)$  si  $TX4(i) = 1$ .

L'échantillon définitif a été scindé en un échantillon principal et un échantillon de réserve.

Seul l'échantillon principal a été mis en production. Il a été intégralement enquêté. Si en cours de collecte l'échantillon principal n'avait pas suffi pour atteindre les cibles souhaitées sur le nombre de répondants, tout ou partie de l'échantillon de réserve aurait été libéré et aurait dû alors être intégralement enquêté.

La constitution de la réserve permet de conserver le caractère aléatoire de l'échantillon, nécessaire pour calculer des intervalles de confiance tout en tenant compte de la contrainte budgétaire, pour éviter une explosion des coûts dans le cas où le taux de réponse ou le taux de hors champs seraient plus favorables que les niveaux retenus pour calibrer l'échantillon.

Il faut souligner le fait que la réserve de « Génération 2010 » n'a pas été mobilisée.

### g) Calcul du taux de sondage

L'échantillon principal de l'enquête « Génération 2010 » est de 279 000 individus sur une population de sortants d'environ un million quatre cent mille individus après correction des défauts de couverture. Le taux de sondage est donc d'environ 20 %.

La probabilité de tirage individuelle est finalement égale à la probabilité d'appartenir à l'échantillon principal ( $TX3(i)$ ) puisque la réserve n'a pas été utilisée.

---

<sup>9</sup> Pour plus d'informations, consulter : [http://www.insee.fr/fr/methodes/outils/cube/documentation\\_cube\\_web.pdf](http://www.insee.fr/fr/methodes/outils/cube/documentation_cube_web.pdf)

### 3. Bilan et perspectives

#### a) Petit historique des plans de sondage des enquêtes « Génération »

##### **Génération 92**

Il s'agit de la première enquête tous niveaux, réalisée par le Céreq. Il n'y a pas eu de constitution d'une base de sondage unique. Le tirage est stratifié, à probabilités inégales. A l'intérieur de chaque strate, on réalise un tirage aléatoire simple (TAS). Les probabilités de tirage ont été optimisées d'un point de vue théorique (en utilisant la formule d'Ardilly) et économique. L'hétérogénéité des strates a été prise en compte dans le calcul grâce à l'utilisateur de 4 indicateurs (durée d'emploi, chômage, dispersion dans les secteurs).

##### **Génération 98**

Constitution de la première base de sondage, avec quelques trous sur des niveaux particuliers (bac+1, etc.). Le plan de sondage a pour objectif de réaliser 30 000 questionnaires au niveau national, bien « proportionnés ». Un tirage complémentaire visant à réaliser 10 000 questionnaires a pour objectif d'assurer une représentativité régionale. Les premières extensions sont menées.

Une estimation du défaut de couverture, du taux de hors champ et de la déperdition est réalisée. Pour mieux connaître les hors champ et la déperdition selon les strates/régions/niveau/spécialité/type d'établissement (public/privé), un marché expérimental est réalisé par un prestataire.

Une matrice de primo-sortants est ensuite développée.

##### **Génération 2001**

L'aspect régional n'est pas pris en compte au moment du tirage, le taux de déperdition est estimé à partir des résultats de « Génération 98 ».

##### **Génération 2004**

Le plan de sondage de l'enquête « Génération 2004 » a été construit de façon à obtenir environ 70 000 questionnaires, 40 000 questionnaires complets et 30 000 questionnaires allégés, répartis de façon à satisfaire les différentes demandes d'extension et selon un tirage équilibré par grands niveaux de formation et régions.

La base de sondage a été divisée en huit ensembles qui correspondent à 8 champs du plan de sondage. Cette stratification est construite à partir du type d'établissements, les apprentis étant mis à part puisqu'une estimation spécifique de la couverture est disponible pour eux : « rectorats », « universités », « formations agricoles », « Jeunesses et Sport » (établissements concernés par l'extension Sport), « écoles de la Santé et du Social » (établissements concernés par l'extension Santé Social), « IUFM », « écoles du supérieur » et « apprentis ». Cette dernière strate comprend tous les apprentis, qu'ils soient sortants de CFA ou non. Les autres regroupements s'entendent donc hors apprentis. Par exemple, le champ « Université » regroupe les sortants des universités hors apprentis. Le tirage de l'échantillon a été effectué séparément pour chacun de ces huit champs.

Pour la première fois, un tirage à deux degrés a été mis en place afin de permettre, en aval, des analyses multi-niveaux, dans le but de voir l'existence possible d'un effet lié à l'établissement de formation dans la vitesse d'insertion des jeunes sortis de ces établissements. Dans les champs pour lesquels cela était possible, des établissements ont d'abord été tirés puis des sortants de ces établissements. Sont concernés les collèges et lycées couverts par le champ Rectorat et les CFA. Dans les six autres champs, il n'y a pas eu de tirage préalable d'établissements. Au final, le tirage à deux degrés concerne 149 000 individus sur les 403 000 de l'échantillon (soit environ 37 %).

*In fine*, le tirage est donc stratifié, à deux degrés sur les champs qui le permettent, et à probabilités inégales. La figure 9 présente ses principales caractéristiques.

Le tirage a été réalisé à l'aide de la macro procédure « Cube » développée par l'Insee. Cet algorithme d'échantillonnage permet d'améliorer la précision des estimateurs associés aux variables d'intérêt de l'enquête. Pour cela, Cube permet de tirer de manière aléatoire un échantillon équilibré sur un ensemble de totaux connus à partir d'informations auxiliaires disponibles dans la base de sondage : la méthode consiste à choisir un échantillon tel que les estimateurs d'Horvitz-Thompson des totaux des variables choisies pour l'équilibrage coïncident avec les vrais totaux. Cette procédure s'applique sur des bases de sondage munies d'informations auxiliaires, qualitatives ou quantitatives, connues au niveau individuel. Des probabilités d'inclusion individuelles destinées à surreprésenter les individus inclus dans des champs d'extensions ont ainsi pu être intégrées dans la procédure de tirage.

Pour l'enquête « Génération », l'équilibrage initial, avant le calcul des coefficients d'extension d'échantillon, a été fait de façon à respecter la répartition des effectifs selon les niveaux de sortie, affinée pour les formations professionnelles du champ. Pour plus d'informations sur le bilan de collecte générale de l'enquête « Génération 2004 », vous pouvez consulter le net.doc sur le site du Céreq à l'adresse :

<http://www.cereq.fr/index.php/publications/Net.Doc/Enquete-Generation-2004-methodologie-et-bilan>

### **Génération 2007**

Le plan de sondage de l'enquête est un tirage aléatoire à un seul degré avec probabilités inégales. La probabilité de tirage individuelle est calculée de façon itérative : un taux de sondage initial permet d'abord de simuler un échantillon répondant aux seuls besoins du Céreq (réalisation de 3 000 questionnaires interprétables sur 10 niveaux de plus haut diplôme). Des surpondérations sont ensuite estimées de façon à répondre aux besoins des différentes extensions d'échantillon (pour atteindre un objectif global de 30 000 questionnaires). Une fois ces calculs réalisés, les poids de sondage finaux calculés permettent de procéder au tirage en une seule étape. Une variable agrégée de strates de formation (Strate07) a été construite afin de faciliter le tirage. Pour plus d'information sur l'enquête « Génération 2007 », vous pouvez consulter le net.doc à l'adresse :

<http://www.cereq.fr/index.php/publications/Net.Doc/Enquete-2010-aupres-de-la-Generation-2007-Bilan-des-redressements>

### **b) L'utilisation de la macro de tirage CUBE appliquée à notre plan de sondage**

Le choix de tirer un échantillon avec des probabilités d'inclusion inégales permet de rendre compte le plus finement possible de l'ensemble des paramètres (taux de couverture, taux de réponse attendu selon les caractéristiques individuelles, gestion des extensions...). La mise en œuvre de la macro CUBE permet de gérer les probabilités individuelles et d'envisager un tirage équilibré à l'intérieur de chaque strate de tirage.

## **IV. Préparation de la collecte**

### **1. Développement du cati (enquête téléphonique)**

#### **a) Logiciel utilisé**

Le questionnaire CATI a été développé sous interviewer Voxco ©. La version utilisée était une version récente orientée web.

Ce questionnaire a pu être testé en ligne, avec une fluidité moindre qu'une version « en dur ».

#### **b) Technique de développement du calendrier**

Le développement du calendrier sous forme de patch (pour éviter la multiplication de la programmation des différents modules par le nombre de mois) n'a pas été choisi.

Le calendrier a donc été développé sous un format linéaire. Toutes les variables relatives à la description d'une séquence étaient dupliquées 45 fois (cas extrême où l'individu changerait de séquences tous les mois) et ce quel que soit le nombre de séquences dans son parcours. Le nombre de variables présentes dans ce type de développement s'est donc élevé à plus de 30 000 et a réduit la vitesse de chargement du questionnaire au moment de la passation. A l'avenir, il paraît évident et indispensable que le développement du calendrier soit systématiquement dynamique, autrement dit que le nombre de modules se « crée » en fonction des différentes situations déclarées.

#### **c) Test du cati et calendrier associé**

Trois tests en réel du CATI ont été programmés au lieu de 2 habituellement.

Date	Programmation	Retour Céreq / Proposition LH2	Solution adoptée
Au 12/11	Calendrier programmé sous forme de questions avec menu déroulant	<b>Retour Céreq</b> : Proposition bloquante et trop éloignée en termes de format, de visuel et de fluidité,	- Affichage d'un visuel de calendrier à la fin de chaque période saisie, - Possibilité de saisie en ordre inversé via une question filtre avant la calendrier,
Au 6/12	Ajout d'item sous le visuel du calendrier pour faciliter le retour arrière sur chaque période saisie	<b>Retour Céreq</b> : Compléter les informations présentes dans les items avec la situation, la date de début et la date de fin de chaque période	Items complétés
Au 30 / 01	Programmation trop lourde lors du pilote 1 (Questionnaire ne supportant pas plus de 8 enquêteurs en simultané dans la matrice du calendrier)	<b>Proposition LH2</b> : Ne garder que le visuel du calendrier de fin, supprimer les visuels intermédiaires après chaque période.	- Alléger la programmation pour qu'elle soit moins basée sur des questions calculées - Sortir les questions calculées de la matrice du calendrier
Au 06/02	Suppression des visuels des calendriers intermédiaires (engendrant un temps de chargement trop long), Optimisation des questions calculées et des masques de saisie	<b>Proposition LH2</b> : Dérouler du questionnaire avec une trentaine d'enquêteur en simultanée. <b>Demande Céreq</b> : Prévoir un test avec 80 enquêteurs en simultané pour dérouler le questionnaire	
Au 12/02	Test en réel avec l'ensemble des enquêteurs présents sur le plateau ( 73 enquêteurs)		
Au 19//02	Suite aux différents tests du questionnaire, un problème de retour arrière dans le calendrier est identifié .		Définition d'une section à la fin de la saisie du calendrier empêchant tout retour arrière.
Au 27/02	Suite au pilote 1b, le calendrier est validé, opérationnel avec 40 enquêteurs en simultané	<b>Retour Céreq</b> : Les 2 premiers pilotes n'ont pas permis de tester le questionnaire en lui-même mais uniquement les capacités techniques.	Un troisième pilote est programmé du 18 au 20 mars (avec 40 enquêteurs)
Jusqu'en Juin	Modifications légères apportées dans le questionnaire		

## 2. Restructuration Normalisation Validation Postale des adresses, et rachat des déménagés

### Renormalisation, validation, postale (RNVP) :

Préalablement à l'envoi des lettres-avis courrier ou mail, une mise à jour des adresses postales a été effectuée de manière à disposer d'adresses fiables, d'une part pour cet envoi, mais également pour les recherches de téléphones en amont de l'enquête, recherches qui sont basées sur l'adresse des individus.

Les recherches par la POSTE sur notre fichier initial de répondants ont donné les résultats suivants :

**Tableau 16 - Résultats RNVP**

	Echantillon principal + réserve	%
<b>Adresses valides</b>	253 449	76 %
<b>Adresses litigieuses</b>	54 538	16 %
<b>Adresses rejetées</b>	25 516	8 %
<b>Total</b>	333 503	100%

- Adresses valides : adresses corrigées et enrichies par la POSTE et réintégrables dans la base
- Adresses litigieuses : doute sur l'adresse
- Adresses en rejet : concernent des voies inconnues dans la localité, déménagés dont l'adresse n'est pas commercialisable (fichier CNIL), etc.



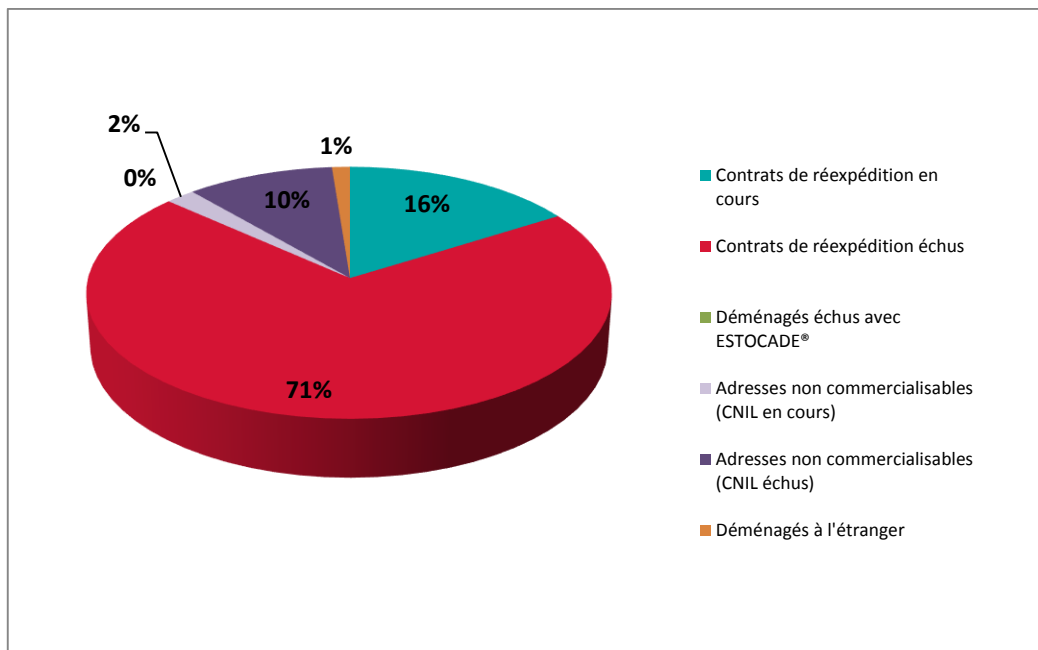
Recherche des déménagés

Sur les 333 503 adresses envoyées, 37 868 individus ont déménagé dont 32 792 sont exploitables (soit près de 10 % des adresses).

**Tableau 17 - Résultats des recherches des déménagés**

Nouvelles adresses	Quantité	% total déménagé	% total adresse
Contrats de réexpédition en cours	6 107	16,13 %	1,83 %
Contrats de réexpédition échus	26 685	70,47 %	8,00 %
<b>Total adresses rachetées</b>	<b>32 792</b>	<b>86,60 %</b>	<b>9,83 %</b>

Topage des déménagés	Quantité	% total déménagé	% total adresse
Déménagés échus avec ESTOCADE®	0	0,00 %	0,00 %
Adresses non commercialisables (CNIL en cours)	771	2,04 %	0,23 %
Adresses non commercialisables (CNIL échus)	3 839	10,14 %	1,15 %
Déménagés à l'étranger	466	1,23 %	0,14 %
<b>Total adresses topées</b>	<b>5 076</b>	<b>13,40 %</b>	<b>1,52 %</b>

**Figure 9 - Répartition des déménagés**

Source : Mediapost

Quelques définitions :

- **Contrats de réexpédition en cours** : nouvelles adresses de déménagés dont le contrat de réexpédition souscrit auprès de La Poste est en cours (pendant 6 mois renouvelable 1 fois) ;
- **Contrats de réexpédition échus** : nouvelles adresses de déménagés dont le contrat de réexpédition souscrit auprès de La Poste est échu ;

- **Adresses non commercialisables CNIL** : Topage anciennes adresses des déménagés qui n'ont pas donné l'autorisation à La Poste de commercialiser leurs nouvelles adresses ;
- **Déménagés à l'étranger** : Topage anciennes adresses de déménagés à l'étranger ;
- **Estocade** : Topage des déménagés dont le contrat de réexpédition est échoué.

### 3. Location de mails auprès de la poste

Pour la collecte des fichiers d'élèves auprès des établissements scolaires, le mail a été collecté en vue d'un envoi de lettre avis par mail.

Dans notre échantillon principal, une sélection a été faite. Les individus n'ayant pas de mail dans la base de sondage ont été recherchés (soit pour 204 482 individus environ).

**Tableau 18 : Résultats des recherches de mails**

Recherche de mail	Echantillon principal	Réserve d'échantillon	Total général
Non	75 002	54 019	129 021
Oui	204 482		204 482
<b>Total général</b>	279 484	54 019	333 503

Sur cette recherche, près de 30 000 mails ont été loués.

### 4. Premier enrichissement des coordonnées téléphoniques

#### c) Le protocole de recherche

La mise à jour des adresses postales permet de disposer d'adresses fiables, d'une part pour l'envoi de la lettre avis, mais également pour les recherches de téléphones en amont de l'enquête, recherches qui sont basées sur l'adresse des individus.

La base de sondage comportait des numéros de téléphones dans 82 % des cas. L'enrichissement de l'échantillon par les coordonnées téléphoniques avait pour triple objectif de compléter les lacunes de la base (faibles cela dit), d'actualiser le cas échéant le numéro de téléphone des personnes qui avaient déménagé et enfin de trouver, le cas échéant, le numéro de téléphone d'un proche de la famille de l'individu ou de trouver d'autres numéros de téléphone pour ceux qui en avaient déjà un.

Il s'agissait donc d'augmenter les chances de joindre les individus à enquêter.

Cette recherche a été confiée après appel d'offre à une société externe disposant des informations présentes dans l'annuaire de France Télécom, mais également d'une base de données contenant des informations sur les numéros de téléphones mobiles et numéros spéciaux comme les lignes dégroupées.

De plus, le prestataire en charge de l'enquête proposait aussi une méthode d'enrichissement de coordonnées téléphoniques via une autre société.

Nous avons donc choisi de coupler les deux afin d'augmenter encore les chances de contact d'individus.

Un protocole de recherche par étape a été retenu.

- **Phase A Prestataire 1**

Dans une première phase, dite « phase A », l'objectif était de retrouver le jeune lui-même ou à défaut ses parents, en recherchant les individus pour lesquels les noms, prénoms et adresses aboutissaient à un seul écho, en acceptant un relâchement léger des critères de recherche : légers écarts d'orthographe sur le nom, le prénom ou l'adresse, voire en l'absence de l'adresse ou du prénom.

- **Phase A prestataire 2**

Cette phase est identique à la première phase de notre premier prestataire. Une comparaison des résultats avait pour objectif de compléter notre première recherche si d'autres numéros étaient trouvés en complément de notre phase A prestataire 1 ;

- **Phase D prestataire 1**

L'objectif était de retrouver le jeune lui-même à partir de cette base partenaire, en excluant les individus retrouvés en phase A ;

- **Phase B prestataire 1**

Relâchement le critère géographique pour rechercher des échos sur les critères du « nom et prénom et commune », « nom et prénom et département », « nom et prénom et région », « nom et prénom et Ile-de-France », « nom et prénom et reste de la France », en autorisant là encore de légers écarts d'orthographe ;

- **Phase C prestataire 1**

Suppression de tout recours au prénom pour rechercher des échos sur les critères « nom et commune » et « nom et département de l'adresse ».

L'ordonnancement de ces différentes étapes s'est fait comme suit :

**Phase A :**

1. NOM PRENOM COMMUNE France télécom si présence d'un à 5 échos

**Sinon Phase D :**

1. NOM PRENOM COMMUNE bases partenaires si présence d'un à 5 échos

**Sinon (phase B et C) :**

2. NOM COMMUNE France télécom et NOM PRENOM DEPARTEMENT France télécom si NOM COMMUNE + NOM PRENOM DEPARTEMENT compris entre 1 et 5 échos (PHASE B)
3. NOM COMMUNE France télécom si présence d'un à 5 échos
4. NOM PRENOM DEPARTEMENT France télécom si présence d'un à 5 échos
5. NOM COMMUNE France télécom et NOM PRENOM DEPARTEMENT bases partenaires si NOM COMMUNE + NOM PRENOM DEPARTEMENT compris entre 1 et 5 échos
6. NOM COMMUNE base partenaire si présence d'un à 5 échos
7. NOM PRENOM DEPARTEMENT base partenaire si présence d'un à 5 échos
8. NOM DEPARTEMENT France télécom si présence d'un à 5 échos
9. NOM DEPARTEMENT base partenaire si présence d'un à 5 échos
10. NOM PRENOM REGION France télécom si présence d'un à 5 échos
11. NOM PRENOM REGION base partenaire si présence d'un à 5 échos
12. NOM PRENOM IDF France télécom si présence d'un écho
13. NOM PRENOM RESTE FRANCE base partenaire si présence d'un écho

## d) Bilan de l'enrichissement

### La phase A prestataire 1 :

Sur les 333 503 individus (échantillon + réserve) envoyés en recherche téléphonique, 47 761 individus ont été retournés en écho des recherches phase A (14 %), c'est-à-dire que pour ces individus on retrouve un écho unique suite à une recherche sur le nom (éventuellement le prénom) et l'adresse de l'individu.

**Tableau 19 - Bilan de la phase A (prestataire 1)**

Classe	Type de résultat	Nombre	%
U11	Nom, prénom et adresse identiques	1 861	0,56 %
U12	Nom et prénom identiques, adresse approchée, une seule réponse	164	0,05 %
U21	Nom ou prénom approché, adresse identique, une seule réponse	1 215	0,36 %
U22	Nom ou prénom approché, adresse approchée, une seule réponse	71	0,02 %
U13	Nom et prénom identiques, adresse absente, une seule réponse	152	0,05 %
O111	Nom identique, adresse identique, prénom différent, une seule réponse	26 841	8,05 %
O131	Nom identique, adresse absente, une seule réponse	1 427	0,43 %
O121	Nom identique, adresse approchée, une seule réponse	3 282	0,98 %
O11	Nom identique, adresse identique, prénom différent ou approché, plusieurs réponses	9 996	3,00 %
O12	Nom identique, adresse approchée, plusieurs réponses	1 642	0,49 %
O13	Nom identique, adresse absente, plusieurs réponses	88	0,03 %
O211	Nom approché, adresse identique, une seule réponse	300	0,09 %
O101	Nom identique, prénom identique, adresse différente, une seule réponse	722	0,22 %
<b>TOTAL</b>		<b>47 761</b>	<b>14,32 %</b>

**Tableau 20 - Bilan de la phase D prestataire 1**

Classe	Type de résultat	Nombre	%
<b>U11</b>	Nom, prénom et adresse identiques	30 076	9,02 %
<b>U12</b>	Nom et prénom identiques, adresse approchée, une seule réponse	1 385	0,42 %
<b>U21</b>	Nom ou prénom approché, adresse identique, une seule réponse	5 711	1,71 %
<b>U22</b>	Nom ou prénom approché, adresse approchée, une seule réponse	243	0,07 %
<b>U13</b>	Nom et prénom identiques, adresse absente, une seule réponse	983	0,29 %
<b>O111</b>	Nom identique, adresse identique, prénom différent, une seule réponse	41 761	12,52 %
<b>O131</b>	Nom identique, adresse absente, une seule réponse	2 183	0,65 %
<b>O121</b>	Nom identique, adresse approchée, une seule réponse	2 547	0,76 %
<b>O11</b>	Nom identique, adresse identique, prénom différent ou approché, plusieurs réponses	9 649	2,89 %
<b>O12</b>	Nom identique, adresse approchée, plusieurs réponses	975	0,29 %
<b>O13</b>	Nom identique, adresse absente, plusieurs réponses	120	0,04 %
<b>O211</b>	Nom approché, adresse identique, une seule réponse	752	0,23 %
<b>O101</b>	Nom identique, prénom identique, adresse différente, une seule réponse	4 847	1,45 %
<b>TOTAL</b>		<b>101 232</b>	<b>30,35 %</b>

*(Recherches excluant les individus retrouvés en phase A)*

Cette phase permet de retrouver 101 232 individus.

### **La phase A prestataire 2 :**

Un total de 34 814 numéros de téléphone a été retrouvé via ce prestataire. C'est quasiment 13 000 numéros de téléphones en moins que l'autre société. A ce stade, la qualité des numéros n'est pas vérifiable, une étude complémentaire pourra être effectuée sur ce point.

### **Bilan des recherches phase B et C (recherches élargies)**

Il s'agit des recherches sur :

- Phase B : le nom + prénom et la commune/le département/la région/Ile-de-France/le reste de la France (homonymes)
- Phase C : le nom et la commune/le département (membres de la famille)

**Tableau 21 - Bilan de la phase B et C prestataire 1**

	Fréquence	%
rech_nomprenom_commune_bp	281	0,3 %
rech_nomprenom_dept_bp	5 433	6,0 %
rech_nom_commune_bp	17 443	19,1 %
rech_nom_dept_bp	13 164	14,4 %
rech_nomprenom_commune_ft	66	0,1 %
rech_nomprenom_dept_ft	2 132	2,3 %
rech_nomprenom_idf_bp	0	0,0 %
rech_nomprenom_idf_ft	1 182	1,3 %
rech_nomprenom_region_ft	809	0,9 %
rech_nomprenom_reste_ft	2 549	2,8 %
rech_nom_commune_ft	26 503	29,00 %
rech_nom_dept_ft	21 700	23,8 %
<b>Total</b>	<b>91 262</b>	<b>100,00 %</b>

91 262 numéros de téléphones ont été sélectionnés pour être remontés.

Globalement :

**Tableau 22 - Bilan chiffré de l'enrichissement**

Nombre de numéros	Nombre d'individus	% d'individus concernés
0	27 818	10,0
1	80 676	28,9
2	94 464	33,8
3	48 572	17,4
4	15 147	5,4
5	7 511	2,7
6	4 085	1,5
7	1 161	0,4
8	50	0,0

Pour 27 818 individus, nous ne récupérons aucun numéro. Pour 80 % des individus, entre 1 et 3 numéros ont été récupérés.

## 5. Ordonnement des numéros de téléphones

Plusieurs numéros sont disponibles pour un même individu. Nous savons que certains numéros sont plus fiables que d'autres, et l'ordre d'appel a donc son importance.

Le choix du numéro selon le type de recherches a été choisi de la manière suivante :

1. Numéro de portable BS confirmé phase D
2. Numéro de fixe BS confirmé phase A
3. Numéro phase D
4. Numéro phase A
5. Numéro de portable BS
6. Numéro de fixe BS
7. Numéro issu de la recherche élargie

## 6. La lettre avis

Cette lettre avis a plusieurs fonctions. Elle a d'abord un rôle informatif. Elle cherche ensuite à sensibiliser le futur enquêté pour faciliter son adhésion lors du contact téléphonique. Elle l'incite à préparer le calendrier d'activité des trois années écoulées, élément clé du questionnaire nécessitant un effort de mémoire. Elle propose également un numéro vert, une adresse courriel et un site internet pour que l'individu puisse indiquer ses nouvelles coordonnées et/ou puisse prendre rendez-vous.

Pour cette génération 2010, une part des individus dispose en plus des informations habituelles, d'une adresse mail.

Pour une question de coûts et de rapidité, le mail a été privilégié pour eux pour l'envoi des lettres avis par rapport à l'envoi de courrier. L'impact de cette procédure (sur les taux de réponse notamment) n'a pas été mesuré.

Les deux modèles de lettre avis figurent dans « Annexe 2 – Lettres avis de contact avec les jeunes ».

Pour tous les individus disposant d'un mail, la lettre avis a été envoyée par mail. Pour les individus restants, et dont les adresses paraissaient suffisamment fiables une lettre-avis a été envoyée par courrier postal dès lors qu'ils n'appartenaient pas à une strate de formation à fort taux de hors champ.

**Tableau 23 - Répartition du nombre de lettres avis envoyées selon le mode et la source**

	Effectif
Lettre papier	103 664
Lettre mail La poste	29 446
Lettre mail Cereq	75 027
Pas d'envoi	71 297
<b>Total</b>	<b>279 434</b>

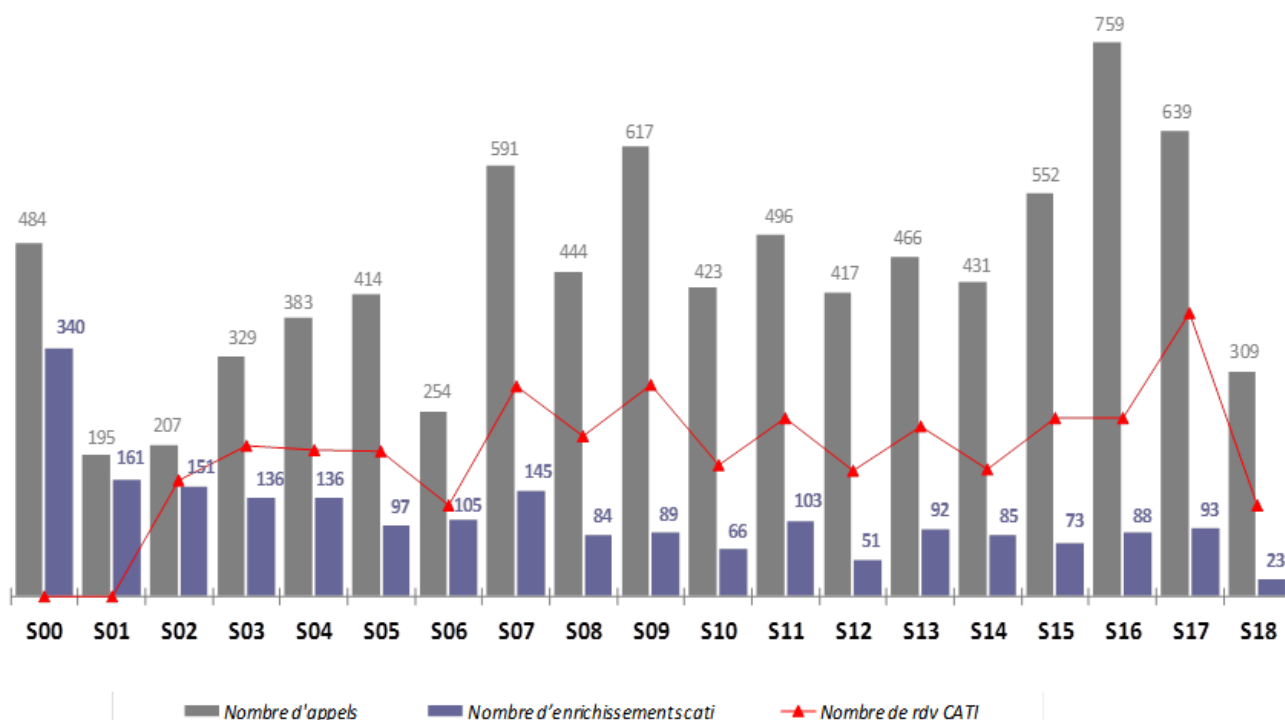
## 7. Le cawi de préfidélisation et le numéro vert

Un site internet dédié aux personnes échantillonnées a été mis en place pour informer sur l'enquête en cours. Accessible à partir de la page d'accueil du site du Céreq, il permettait aux jeunes d'avoir des informations sur l'enquête ainsi que de consulter les résultats des enquêtes précédentes, mais surtout d'accéder à un espace internet (appelé CAWI de préfidélisation ou d'enrichissement) afin d'y enrichir ses coordonnées téléphoniques et/ou de prendre rendez-vous.

De la même manière, un numéro vert a été mis en place pour permettre aux jeunes de donner de nouvelles coordonnées, de réaliser l'enquête ou de faire part de leur refus d'y répondre. Le numéro était présent sur la lettre-avis. Il était aussi communiqué lors des contacts téléphoniques à des tiers qui refusaient de transmettre les coordonnées de l'individu ou aux proches d'individus difficiles à joindre.

Ce numéro vert a été ouvert du lundi au samedi, entre le 25 mars et le 31 juillet 2010. En dehors des horaires d'ouverture, les jeunes pouvaient laisser un message sur un répondeur. Il a reçu 8 410 appels.

Figure 10 - Enrichissements des coordonnées via le numéro vert



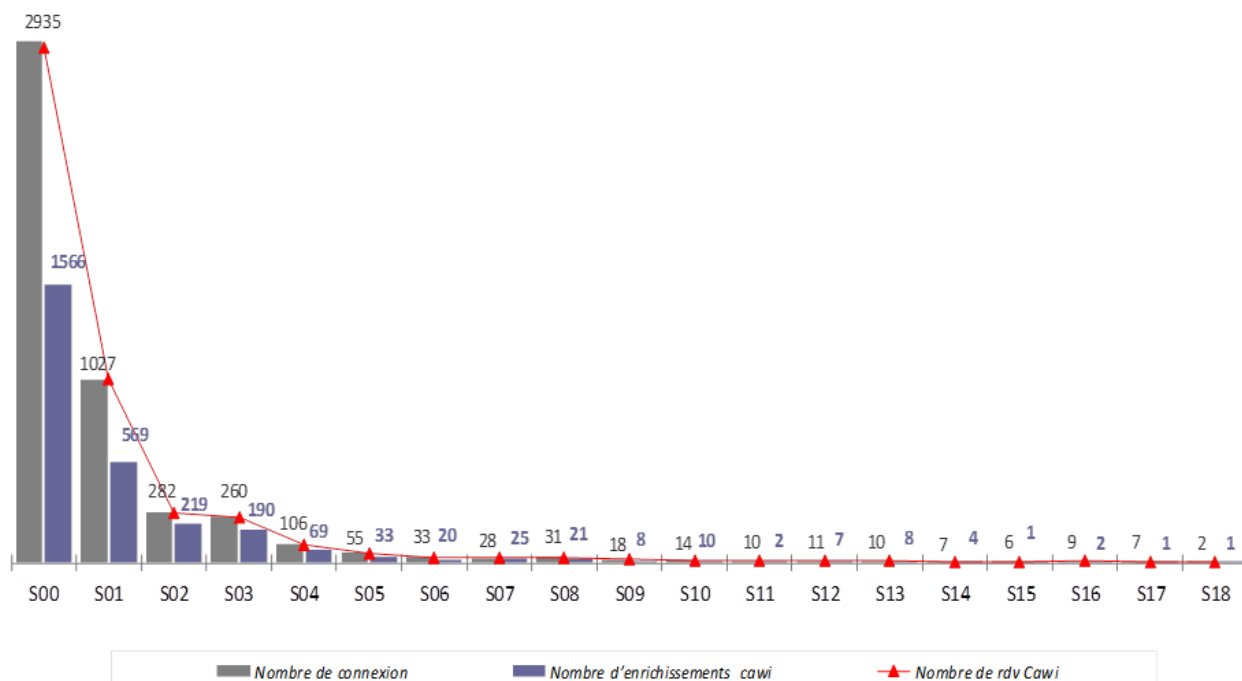
Source LH2

S00 = dernière semaine de mars

S01 = première semaine d'avril (= démarrage de la collecte)

S18 = dernière semaine de juillet (= fin de collecte)

Figure 11 - Enrichissement des coordonnées via le Cawi



Source LH2



## 8. Une recherche supplémentaire de numéros de téléphone en cours de plateau

Malgré les procédures d'enrichissement en numéros de téléphone assez complètes, nous nous sommes aperçus que certains numéros présents dans les pages blanches n'étaient pas remontés par les prestataires.

Une recherche complémentaire via un système automatisé d'interrogation des pages blanches a été mis en place durant le plateau d'enquête, afin de compléter les numéros déjà présents dans les bases.

## 9. La phase de pré qualification

Cette phase a été « transparente », les questions dédiées ont été directement intégrées au questionnaire CATI pour passer le questionnaire directement après la phase de pré-qualification.

## 10. Relations et contacts avec les enquêtés

L'équipe en charge de l'enquête « Génération » a imaginé d'autres moyens de fidéliser la population de jeunes répondants à l'enquête, autres que la lettre-avis, notamment dans le but d'amortir les effets de l'attrition pour les futures ré-interrogations.

Dans cette optique, communiquer à travers les réseaux sociaux ou par l'intermédiaire de courrier électronique, permet de développer une réelle proximité à la différence du courrier papier. Ce contact est activé avant l'enquête mais dorénavant en cours de plateau et post-enquête. Les moyens de communication privilégiés ont été adaptés au profil jeune des répondants.

### a) Site internet du Céreq

L'échantillon de l'enquête « Génération 2010 » est contacté une première fois à l'aide d'une lettre-avis (papier ou mail) qui informe des modalités de contact ultérieur, des thématiques abordées dans le questionnaire et de la durée du plateau d'enquête. Pour permettre aux jeunes de s'assurer de l'origine et de l'intérêt de l'enquête, des informations ont été diffusées sur le site du Céreq ([www.cereq.fr](http://www.cereq.fr)). Dès la page d'accueil, un encart visible sur toute la durée du plateau d'enquête et une page spécifique « Génération 2010 » précisent les objectifs de l'enquête et l'importance de leur participation.

### b) Réseaux sociaux : Facebook et Twitter

Les plateaux d'enquête de « Génération » durent trois mois environ durant lesquels les jeunes tentent de s'informer sur l'enquête avant de répondre à nos appels. Au-delà du site internet du Céreq, une page Facebook « Céreq « Génération 2010 » ([www.facebook.com/cereqG10](http://www.facebook.com/cereqG10)) a été créée pour pallier ce manque d'interactions et d'échanges avec l'équipe en charge de l'enquête. De fait, l'alimentation de ce compte est réalisée par l'équipe.

La création de la page Facebook répond à plusieurs objectifs :

- créer un lien entre le Céreq et les jeunes afin de les fidéliser pour les prochaines interrogations (à 5 et 7 ans) ;
- recueillir les commentaires des jeunes sur l'enquête ;
- donner plus de visibilité au Céreq et à l'enquête « Génération » en particulier.

Au-delà de l'information donnée sur le Céreq et ses missions, la page Facebook était principalement destinée à informer les jeunes répondants sur l'enquête « Génération » elle-même. La page contenait

les objectifs de l'enquête, les modalités, le déroulement, la périodicité et les dates d'interrogation, la présentation de l'équipe en charge de l'enquête, les actualités, les événements en lien avec l'enquête, les informations post-enquêtes, des vidéos et les premiers résultats. Toutes les informations qui permettent de favoriser la réponse positive à l'enquête.

Par exemple, répondre au questionnaire permet d'enrichir les informations sur leur orientation, sur le choix des filières, la situation économique à l'entrée sur le marché du travail, le type de contrat proposé, le salaire, etc.

La page Facebook a par ailleurs permis d'ouvrir les portes du plateau d'enquête ce qui a facilité le premier contact avec les télénquêteurs. Une photo du plateau d'enquête et des télénquêteurs a été publiée sur la page.

Audience Facebook :

- 383 fans (condition : avoir un compte Facebook personnel) ;
- plus de 2 000 visiteurs/par semaine (ayant interagi ou non avec la page) ;
- profil des fans/visiteurs : majoritairement des femmes de 25-34 ans.

Post le + populaire :

Génération 2010 - CEREQ a posé une question.  
20 juin

Et vous ? Après 3 ans de vie active, recherchez-vous ou occupez-vous un emploi en rapport avec votre formation initiale ?

<input type="radio"/>	Oui, totalement	+26
<input type="radio"/>	Oui, partiellement	+11
<input type="radio"/>	Non	+13

Je n'aime plus · Commenter · Partager 3 1

1 608 personnes ont vu cette publication Stimuler la publication

Le sentiment des jeunes interrogés :

Des commentaires positifs et de nombreux encouragements. Les participants ont trouvé que la démarche était utile, que leur participation était valorisante. Certains ont signalé que le questionnaire était long mais pour la bonne cause. Et d'autres ont indiqué que le calendrier professionnel leur a permis de faire une rétrospection de leur parcours.

Ce qui leur a paru négatif relève de la longueur de l'entretien, de la répétition de certaines questions et de l'inadéquation avec certaines professions « il manque des cases ». Enfin, plusieurs d'entre eux regrettaient de n'avoir la possibilité de s'exprimer plus.

Expérience plutôt positive :

La page Facebook a rendu l'enquête plus accessible tout en lui conférant un caractère officiel. Cet outil de communication est adapté au public cible avec en plus le côté interactif. Beaucoup de commentaires utiles pour améliorer la qualité de la collecte des enquêtes futures (notamment sur la qualité de l'expérimentation internet).

Il existe un compte Twitter du Céreq (<https://twitter.com/PRESSECEREQ>), orienté presse, destiné plus particulièrement aux journalistes et au réseau institutionnel du Céreq.

### c) Mailings

Pour fidéliser les jeunes répondants, des messages électroniques ont été envoyés à chacun des participants. Pour cela, l'adresse mail a été collectée à la fin du questionnaire et une adresse de messagerie de contact a été mise en place pour faciliter les échanges ([generation2010@cereq.fr](mailto:generation2010@cereq.fr)).

Créer le lien passe par la connaissance de l'enquête mais aussi de celui qui la produit : le Céreq. Chaque mail contient un lien vers le site internet du Céreq mais aussi l'adresse de la page Facebook. Le mailing a permis notamment de vérifier au fil de l'eau la validité des mails recueillis.

Deux mailings ont été réalisés :

1. Un mailing de remerciement au cours du plateau d'enquête : il a été envoyé en 4 paquets distincts correspondant aux quatre mois d'interrogation (prévu en fin de mois)

	Avril	Mai	Juin	Juillet	Ensemble
Nombre de mails envoyés	7 685	9 754	9 860	3 312	30 611
Echec (confondus)	244	620	825	315	2 004

Le taux d'échec est aux alentours de 7 %. Les échecs sont pour la plupart des mails qui ont été spammés mais aussi des mails erronés, des boîtes de réception pleines, ...

Mail type de remerciement :

Bonjour <Prénom> <NOM>,

Nous tenons à vous remercier chaleureusement pour votre participation à l'enquête « Génération 2010 » à laquelle vous avez répondu par téléphone au cours du mois de <mois d'interrogation>.

Si vous souhaitez vous exprimer sur l'enquête, utilisez sa page Facebook : [www.facebook.com/cereqG10](http://www.facebook.com/cereqG10). Vous retrouverez des informations sur les enquêtes « Génération » en cliquant [ici](#).

*L'équipe gestion d'enquêtes du Céreq.*

2. Un mailing d'information contenant les premiers résultats de l'enquête : envoyé au moment de la sortie du Bref « Face à la Crise, le fossé se creuse entre niveaux de diplômés », en avril 2014.

Mail type d'envoi des résultats :

Bonjour <Prénom> <NOM>,

Le Céreq vous a interrogé en <mois d'interrogation> 2013 dans le cadre de son enquête sur l'insertion professionnelle des jeunes de la « Génération 2010 ».

Nous avons le plaisir de vous transmettre [les premiers résultats](#) et vous invitons à réagir sur cette étude sur la page [Facebook](#) de l'enquête.

Nous vous remercions d'avoir contribué à faire progresser la connaissance sur les mécanismes d'insertion des jeunes sur le marché du travail afin d'éclairer les politiques publiques en matière de formation et d'emploi.

*L'équipe gestion d'enquêtes du Céreq.*

Le mailing a été réalisé à l'aide du logiciel Power Emailing 4 avec personnalisation des mails (prénom et nom de l'enquêté en en-tête, ...). Le mailing est ensuite envoyé par paquet de 500 toutes les 5 minutes.

Une mise à jour de la base de mails des répondants de la « Génération 2010 » est en cours.

Certains répondants ont indiqué par retour de mail leurs nouveaux numéros de téléphone, ils seront également intégrés.

## V. La collecte par téléphone

La collecte de l'enquête s'est déroulée par téléphone d'avril à juillet 2013 chez le prestataire LH2 (basé à Saint Ouen - 93400). 279 484 individus ont été mis en production, la réserve tirée n'a pas été utilisée. Le plateau d'enquête a été exceptionnellement prolongé en juillet au vu du nombre insuffisant de répondants (environ 30 000 questionnaires réalisés fin juin).

L'enquête est assistée par ordinateur, donc l'enquêteur interroge les individus par téléphone tout en suivant sur un écran d'ordinateur un script préétabli qui affiche les questions qu'il doit poser et les éventuelles modalités de réponse, avec la consigne de lire strictement les questions et de lister les modalités, sauf mention contraire. Les réponses sont saisies directement sur informatique. Le Cati est capable de gérer des filtres et d'orienter l'individu vers des questions différentes en fonction des réponses précédemment données.

Auparavant, les télé-enquêteurs étaient spécialisés sur des cibles particulières (selon le diplôme, le type de questionnaire, le résultat des recherches téléphoniques), l'échantillon était subdivisé en différents terrains d'enquêtes. Pour la première fois, il n'y a pas eu de spécialisation des enquêteurs, sans que l'on puisse facilement mesurer l'impact sur les taux de réponse, ou encore sur la facilité de compréhension des enquêtés selon l'enquêteur.

### 1. Calendrier et organisation générale de la collecte

Figure 12 - Calendrier et organisation générale avec le prestataire

	Délais
Réunion de démarrage	17 octobre 2012
Transmission du questionnaire	22 octobre 2012
Programmation et tests du questionnaire CATI	Novembre 2012 – Janvier 2013
Transmission au CEREQ d'un lien test	
Transmission lien CAWI d'enrichissement des coordonnées	22 Novembre 2013
Réunion programmation (LH2) – (Céreq)	17 janvier 2013
Transmission Fichier Pilote 1	25 janvier 2013
Premier pilotage CATI – 67 questionnaires complétés – 43 min en moyenne	28 – 30 janvier 2013
Transmission Fichier Pilote 2	25 février 2013
Deuxième pilotage CATI auprès de 95 individus – 34 min en moyenne	26 -27 février 2013
Troisième pilotage CATI auprès de 340 individus – 41,3 min en moyenne	18- 20 mars 2013
Ouverture du numéro vert	25 mars 2013 (Semaine 00)
Formations enquêteurs	2-4 avril (Semaine 01)
Réalisation des enquêtes	Du 2 avril au 31 juillet 2013 (Semaine 01-S18)
Réunion intermédiaire	24 mai 2013
Remise du fichier final CATI	1 août 2013

## 2. Le suivi de la collecte en chiffres

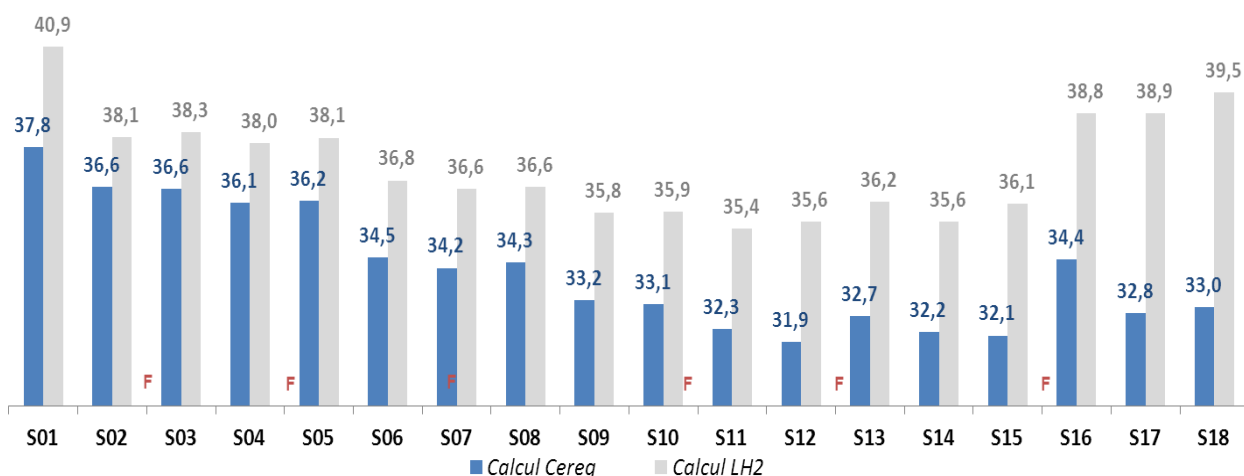
67 % des enquêtes complétées ont été réalisées par portable.

Pour la première fois et à titre expérimental, les individus résidant à l'étranger ont eu la possibilité de répondre à un questionnaire par internet. Ils sont considérés hors champ de l'enquête et feront l'objet d'une analyse ultérieure.

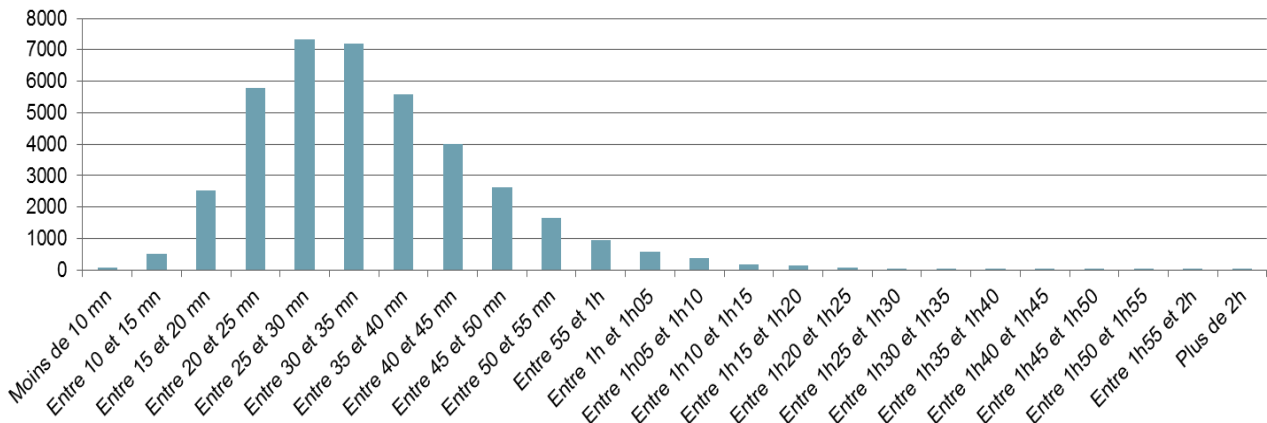
**Figure 13 : Nombre d'enquêtes réalisées selon la période et le nombre de télé-enquêteurs**



**Figure 14 : durée moyenne du questionnaire selon la semaine d'enquête**



**Figure 15 - Distribution de la durée de passation du questionnaire**



Parmi les répondants ayant terminé le questionnaire, 41 % des entretiens ont duré moins de 30 minutes, 10 % ont duré plus de 45 minutes et 3,7 % ont duré plus d’une heure.

En moyenne, la passation du questionnaire s’est déroulée en 34 minutes.

Sur le graphe de l’évolution des durées par semaine d’enquête, il apparaît que le turn-over des télé-enquêteurs a eu un impact sur le temps de passation (notamment en fin de collecte).

Tout au long de la collecte, les membres de l’équipe en charge de l’enquête (avec l’appui de 2 personnes qualité) ont régulièrement supervisé le plateau d’enquête. Ils ont également assuré 14 sessions de formation (296 enquêteurs formés pour 221 opérationnels).

Le prestataire a rencontré des problèmes techniques et divers ralentissements : sur les 56 456 heures de production, il estime ainsi une perte d’environ 1 530 h. Avec un rythme final de 0,7 questionnaire par heure de production, ces heures auraient représenté l’équivalent d’environ 1 070 questionnaires.

### 3. Les règles de rappel

**Tableau 24 - Les règles d’appel détaillées**

Rappel	Résultat	Fréquence	Consécutif	Interviewer	Délai	Unité	Action sur la priorité	Priorité	Compteur + 1
VRAI	03 (INT03) Hors-cible	1	FAUX	Libre	4	Jours	=	2	VRAI
VRAI	10 (INTRO) Faux numéro Disque France-Télécom	1	FAUX	Libre	3	Jours	=	2	VRAI
VRAI	11 (INTRO) NRP	1	FAUX	Libre	1	Quarts	=	2	VRAI
VRAI	12 (INT12) RV	1	FAUX	Libre	0	Minutes	=	5	FAUX
VRAI	13 (INTRO) OCCUPE	1	FAUX	Libre	10	Minutes	=	3	FAUX
VRAI	13 (INTRO) OCCUPE	2	VRAI	Libre	60	Minutes	=	2	FAUX
VRAI	14 (INTRO) REPONDEUR	1	FAUX	Libre	240	Minutes	=	2	FAUX
VRAI	15 (INTRO) Il n'y a pas de <ren> <nom> ici	1	FAUX	Libre	4	Jours	=	2	VRAI
VRAI	16 (INT16) Ne connaît pas le numéro de la cible	1	FAUX	Libre	10	Jours	=	2	VRAI
VRAI	29 (INT29) va rappeler sur le numéro vert	1	FAUX	Libre	5	Jours	=	2	FAUX
VRAI	30 Interview arrêtée - Rappeler (Rappel ferme)	1	FAUX	Libre	0	Minutes	=	5	FAUX
VRAI	60 Problème de système tél.	1	FAUX	Libre	1	Jours	=	2	VRAI
VRAI	84 Rappel immédiat avec saisie d'un nouveau numéro	1	FAUX	Même interviewer que l'appel précédent	5	Minutes	=	2	VRAI
VRAI	85 (INTRO) A raccroché	1	FAUX	Libre	15	Minutes	=	2	FAUX
VRAI	86 (INTRO) Message laissé sur le répondeur	1	FAUX	Libre	1	Jours	=	2	FAUX
VRAI	P1 Pas de réponse	1	FAUX	Libre	1	Quarts	=	2	VRAI
VRAI	P2 Occupé	1	FAUX	Libre	1	Quarts	=	3	VRAI
VRAI	P3 Interception par l'opérateur	1	FAUX	Libre	1	Jours	=	2	VRAI
VRAI	P4 Répondu	1	FAUX	Libre	1	Quarts	=	3	VRAI
VRAI	P5 Répondeur	1	FAUX	Libre	1	Quarts	=	2	VRAI
VRAI	P6 Fax / Modem	1	FAUX	Libre	3	Jours	=	2	VRAI
VRAI	P7 Pas de sonnerie - Pas de tonalité	1	FAUX	Libre	3	Jours	=	1	VRAI
VRAI	P8 Valeur par défaut	1	FAUX	Libre	4	Jours	=	2	VRAI

**Grille de Lecture :**

Une fiche codée Int=03 en J sur TEL01, a été rappelée en J+4 (Délai) sur TEL02 (Compteur+1= vrai).

Une fiche codée Int=13 en J, sera rappelée 10 minutes après, si elle est à nouveau codée Int=13 elle sera rappelée 60 minutes après sur le même téléphone (Compteur+1= faux).

#### 4. Des messages laissés sur répondeur

A partir du 11 juillet, le processus de laisser un message sur répondeur a été lancé. Cette démarche avait pour objectif de sensibiliser les individus qui ne répondaient pas à la vue du numéro qui s'affichait. (Le numéro était visible pour la première fois sur une enquête génération).

##### Encadré 5 - Message laissé sur les répondeurs des interlocuteurs :

Bonjour Madame/Monsieur <PRENOM> <NOM>, <NOM ENQUETEUR> du Centre d'Etudes et de Recherche sur les Qualifications.

Nous réalisons actuellement une étude nationale sur l'insertion des jeunes et votre avis est très important.

Nous avons des difficultés à vous joindre. Contactez-nous au numéro vert suivant 0 800 97 00 55 afin de fixer un rendez-vous pour répondre à notre étude.

Ce numéro vert est gratuit depuis un téléphone fixe. Il est ouvert en semaine de 9h00 à 20h30 et le samedi de 10h00 à 17h30. Dans l'attente de votre appel, merci et très bonne journée.

#### 5. Suivi technique et personnes « qualité »

En raison de la distance entre le plateau d'enquête (Paris Saint-Ouen) et le Céreq, basé à Marseille, deux personnes « qualité » ont été recrutées pour suivre le plateau téléphonique en continu. Cela a notamment permis au Céreq d'avoir de l'information en temps réel. Ces personnes « qualité » ont réalisé des écoutes quotidiennes, ont assisté les superviseurs et assuré la formation des enquêteurs lorsque les agents du Céreq n'étaient pas présents. Elles ont fait part des difficultés rencontrées sur place et permis au Céreq de réagir efficacement face aux problèmes humains, techniques ou d'organisation du terrain. Un bilan de collecte a été réalisé par ces personnes « qualité » : *Rapport sur le déroulement de l'enquête Génération 2010*.

#### 6. Les résidents étrangers à la date d'enquête

Une expérimentation de l'enquête « Génération » en version internet (CAWI) a été testée. Pour la première fois, le comité du label a préconisé, à titre exploratoire, l'interrogation des jeunes sortant du système éducatif en 2009-2010 et résidant à l'étranger à la date de l'enquête (en 2013). Ces jeunes n'étant pas joignables par téléphone, leurs adresses mails ont été récupérées par l'intermédiaire d'un tiers. Ils ont eu la possibilité de répondre au questionnaire par internet. Ces 334 questionnaires ne sont pas disponibles dans les bases, les résidents à l'étranger sont hors champ de l'enquête. Cette expérimentation fera l'objet d'une expertise afin de déterminer la possibilité d'une future intégration dans le champ de ces jeunes dans les enquêtes.



## VI. Les traitements en aval

### 1. Les tables brutes

La structure des tables fournies par le prestataire est directement liée au format de développement du CATI.

Autrefois, nous disposions de trois extractions pour l'enquête « Génération » : une table dite « individus » qui concernait toutes les données individuelles situées avant le calendrier, une table calendrier en ligne pouvant contenir plusieurs lignes pour un même individu, et une table qui contenait les données sur la situation à date d'enquête ainsi que les informations de fin de questionnaire.

Pour cette enquête « Génération 2010 », le CATI n'intégrait pas de questionnement dynamique. Le calendrier a été développé sous format linéaire prévoyant ainsi 45 mois (de novembre 2009 à juillet 2013) pour six situations différentes (intérim, entreprise, recherche d'emploi, formation, reprise d'étude ou autre situation).

Par exemple, notre première question du module d'emploi, la EP1 « vous étiez salarié, aide familiale, à votre compte ? » était donc prévue 45 fois dans le développement du CATI, soit 45 variables nommées de EP1\_1 à EP1\_45 (le numéro d'incrémentation concernant le numéro de la séquence décrite). Si l'individu avait uniquement deux séquences d'emploi dans son parcours, Ep1\_1 et Ep1\_2 étaient renseignés (quelle que soit la durée de la séquence) et toutes les autres étaient vides. Ce fichier non optimisé s'est révélé lourd à gérer par la suite.

Plusieurs extractions nous ont été livrées en cours d'enquête afin de faire des contrôles et des vérifications. Une extraction de données à une date précise permettait de récupérer une table (scindée en trois pour faciliter les imports), contenant la totalité des individus ayant répondu à l'enquête (à une date donnée) et la totalité des variables (34 000 colonnes quel que soit le nombre de répondants).

Nos premières lectures ont été compliquées, car d'une part, le délai d'ouverture d'une table même scindée en trois (d'environ 10 000 colonnes chacune) était très long, qu'un import sans de cette ampleur n'était pas immédiat, et pour finir qu'une lecture rapide était impossible à effectuer en l'état.

Un temps conséquent a été nécessaire pour développer une macro sas optimale pour lire et mettre en forme les données fournies afin qu'elles soient lisibles.

Cette macro et un premier tri de variables nous ont permis d'obtenir deux tables : une table individu (contenant la totalité des variables individuelles) et une table calendrier (contenant la totalité des séquences par individu pour chacun des 6 modules possibles).

Suites aux retraitements habituels, nous avons pu obtenir les trois tables habituelles de notre enquête « Génération », une table « individu », une table « seqentr » (séquences d'emploi), une table « nonempl » (séquences de non emploi : chômage, formation, reprise d'études, autres situations d'inactivité, ...).

## 2. La création des tables

Cette partie présente l'ensemble des traitements réalisés pour construire les fichiers finaux à partir des fichiers bruts, à l'exception du calcul de la pondération finale présentée dans la partie VII, qui aborde le traitement de la non réponse globale et le calage sur marge.

Les fichiers des questionnaires de l'enquête principale CATI et de l'enquête expérimentale internet CAWI ont été traités ensemble dans un premier temps. De plus, les observations entrant dans le champ Céreq et celles relevant d'extensions spécifiques (les post-initiaux des formations du sport, des formations sanitaires et sociales, et des formations des écoles de fonctionnaires) n'ont pas été séparées pour les phases d'apurement, de codification, de création de variables. Elles n'ont été dissociées qu'à la fin, pour la constitution des bases finales et les calculs de pondération.

### a) La construction des bases brutes

Rappelons que le prestataire a livré une table scindée en trois de 38 500 individus et 34 000 colonnes (pour l'enquête principale). Dans un premier temps, ces tables ont été mises en forme pour obtenir une table « Individus » et une table « calendrier » globale.

Pour la construction des premières bases, les informations du calendrier professionnel ont été éclatées en deux parties de façon à construire une table sur les périodes d'emploi et une table sur les périodes de non emploi. Une table sur les individus a par ailleurs été construite pour rassembler les informations du calendrier de mode de cohabitation et celles du reste du questionnaire.

La table « Individus » contient donc les informations sur l'identification de l'enquêté, son parcours scolaire, l'identification du plus haut diplôme, les diplômes obtenus après 2010, les séjours à l'étranger, les stages en cours d'études, ses caractéristiques individuelles et son environnement familial ainsi que toutes les informations demandées dans les modules spécifiques.

La table « Séquences d'emploi » décrit chaque séquence d'emploi à partir du calendrier professionnel. La table « Séquences de non emploi » décrit chaque séquence de non emploi (chômage, formation, reprise d'études, autres situations d'inactivité, ...) à partir du calendrier professionnel.

Une première phase de validation consiste ensuite à vérifier que l'ensemble des variables présentes dans les questionnaires étaient remplies et, pour chaque variable, que les modalités de réponses correspondaient à celles des questionnaires.

Dans la table « Individus », des variables relatives à la trajectoire, aux diplômes, au calendrier professionnel et au calendrier habitat ont été créées.

### b) Les variables de la table « Individus » construites à partir des calendriers

#### Variables issues du calendrier professionnel

Les variables MOIS1 à MOIS45 permettent de suivre mois par mois, de novembre 2009 (mois1) à la date d'interrogation maximale en juillet 2013 (mois45), la situation professionnelle du répondant (elles fournissent également l'identifiant de la période concernée dans la table des périodes d'emploi ou celle des périodes de non emploi).

Les variables ID, IF et DUROBS correspondent respectivement aux numéros du premier et du dernier mois observé et au nombre total de mois renseignés dans le calendrier (DUROBS= IF-ID+1).

Les variables NMEMP, NMJVAC, NMCHO, NMINA, NMFOR, NMETU et NMVAC renseignent sur le nombre total de mois passés dans chaque situation professionnelle, respectivement en emploi, en job de vacances, au chômage, en inactivité, en formation, en études et en vacances.

Les variables NSEMP, NSJVAC, NSCHO, NSINA, NSFOR, NSETU et NSVAC renseignent sur le nombre total de séquences passées dans chaque situation professionnelle. La variable NSTOT donne le nombre total de séquences de l'individu.

La variable TAPE renseigne sur le temps d'accès au premier emploi (si aucun emploi occupé, TAPE est manquante), TRAJPRO correspond à une typologie de parcours et SITDE à la situation professionnelle à la date d'enquête.

#### Variables issues du calendrier du mode de cohabitation

Les variables HMOIS1 à HMOIS45 ont été créées pour donner le mode de cohabitation (seul, en couple, chez ses parents) mois par mois de novembre 2009 (mois1) à la date d'interrogation.

#### Questions à choix multiples :

Certaines questions autorisent les réponses multiples. Après codage des réponses ouvertes, le cas échéant, les différentes réponses possibles ont donné lieu à la création d'autant de variables dichotomiques, chaque modalité de réponse proposée étant associée à une variable booléenne (Oui/Non).

Pour exemple, la variable CA26 (discrimination à l'embauche) question multiple propose 8 réponses possibles, et une réponse ouverte en 9<sup>e</sup> position.

Les réponses ouvertes sont reclassées dans les variables existantes quand c'est possible et une autre modalité peut être créée si elle est pertinente et en nombre suffisant. Dans cet exemple, la CA26 finale comporte 18 modalités. Les 18 variables dichotomisées correspondantes allant de CA26\_1 à CA26\_18 ont donc été créées.

### **c) La vérification des filtres**

Selon leurs caractéristiques et leurs réponses aux questions précédentes, les individus ne doivent pas tous répondre aux mêmes items.

Auparavant, dans le cas où l'enquêteur procédait à des retours arrière lors de la passation du questionnaire, les réponses restaient enregistrées. Dans ce cas, certaines variables qui n'auraient pas dû être renseignées l'étaient. Le travail consistait alors à vérifier les filtres et à supprimer ces réponses erronées.

Pour la première fois, un nettoyage automatique a été paramétré chez notre prestataire pour éviter ces réponses erronées. Les retours-arrières sont donc devenus invisibles dans la table mise à disposition, car déjà traités.

### **d) La suppression d'observations non exploitables**

En théorie, un calendrier d'activité mal renseigné ne pouvait être validé et l'enquête ne pouvait alors pas se poursuivre.

En pratique, il est arrivé dans certains cas, des bugs de calendrier mal définis, laissant certains calendriers d'activité non exploitables, à cause de certaines périodes qui se chevauchent, ou qui ne sont pas renseignées ou encore parce que des périodes d'emploi mentionnées ne sont pas associées à une description de l'emploi alors que cela aurait dû être le cas.

Ces observations jugées non exploitables sur le noyau du questionnaire ont été supprimées.

Par ailleurs, certains répondants se sont révélés hors champ a posteriori, comme des individus qui mentionnent un arrêt des études pendant le questionnaire filtre et qui commencent alors le calendrier professionnel par une reprise d'étude.

Au total, moins de 3 % des observations récupérées ont été supprimées.

### e) Enrichissement par les informations de la base de sondage

La table des « Individus » est enrichie par des informations issues de la base de sondage et relatives à la scolarité suivie en 2009-2010 : formation suivie, niveau, spécialité, ...

Ces informations complètent les réponses au questionnaire sur les informations concernant le parcours scolaire (série du bac, diplômes, etc.).

### f) Les recopies

Les recopies ne concernent que les séquences d'emploi :

- emplois courts inférieur à 6 mois du passé (CAL=01) ou actuels (CAL=03) pour les salariés et intérimaires (INTER=1 ou INTER=2) ;
- emplois longs inférieur ou égal à 12 mois du passé (CAL=02) ou actuels (CAL=04) pour les salariés et intérimaires dans une seule entreprise (INTER=2) **hors** intérimaires plusieurs missions dans plusieurs entreprises différentes (INTER=1).

On distingue plusieurs situations :

- pour les emplois longs du passé ou actuels (CAL=02 ou CAL=04), si l'individu change de situation, c'est à dire un changement de (et/ou) :
  - o contrat de travail ;
  - o profession ;
  - o catégorie socio professionnelle ;
  - o temps de travail ;
  - o salaire, primes et 13<sup>e</sup> mois.

Alors on pose les questions de début et de fin de séquence. Il n'y a pas de recopies. En revanche, s'il indique que l'une de ses situations n'a pas changé, on recopie le début sur la fin :

- pour les emplois courts du passé (CAL=01), on ne pose systématiquement que les questions de début de séquence. Les informations saisies dans les variables de début de séquence qui concernent l'embauche sont recopiées dans les variables de fin de séquence qui concernent le départ de l'entreprise ;
- pour les emplois courts actuels (CAL=03), on ne pose systématiquement que les questions de fin de séquence. Les informations saisies dans les variables de fin de séquence qui concernent le départ de l'entreprise sont recopiées dans les variables de début de séquence qui concernent l'embauche. C'est le seul cas où les informations de fin sont copiées sur le début de séquence.

### g) L'anonymisation des fichiers

Afin d'assurer la confidentialité des réponses des personnes enquêtées et conformément au principe du secret statistique garanti par la loi de n° 51-711 du 7 juin 1951 sur l'obligation, la coordination et le secret en matière de statistiques, les fichiers de résultats ont été rendus anonymes.

L'anonymisation consiste d'abord à supprimer les informations directement nominatives (nom et prénom des personnes), ainsi que les coordonnées téléphoniques et postales, nécessaires à la réalisation de la collecte de l'enquête. Les localisations géographiques trop fines (en l'occurrence la commune), les noms des entreprises sont également supprimés.

Elle consiste aussi à s'assurer qu'une identification indirecte par recoupement des différentes informations disponibles est impossible. Pour cela, certaines variables ont été supprimées, d'autres ont vu leurs modalités regroupées. En particulier, les informations de localisations géographiques fines ne sont pas diffusées (code commune etc.). Les questions ouvertes sont supprimées, seule est conservée la codification de ces variables.

### 3. La codification des variables

#### a) Le repérage du plus haut diplôme obtenu

Le plus haut diplôme est créé à partir d'une hiérarchisation normalisée de l'ensemble des diplômes décrits dans le module parcours scolaire.

Le plus haut diplôme est déterminé à partir des trois informations suivantes :

- le fait d'avoir obtenu le diplôme de sortie pour les élèves en classe terminale en 2009-2010 ;
- le fait de posséder un baccalauréat ;
- la liste des diplômes cités par l'individu lors du parcours scolaire.

Le plus haut diplôme obtenu ne correspond donc pas toujours au niveau de formation de la classe de sortie, par exemple du fait des échecs aux examens et abandons en cours d'études. Pour un petit nombre d'individus, le diplôme de sortie peut également ne pas être le plus haut diplôme, du fait de réorientations vers des formations de niveau inférieur au cours du parcours scolaire.

Avec la « Génération 2010 », une réflexion approfondie a été menée dans un groupe de travail pour améliorer le repérage du plus haut diplôme obtenu. Les questions du module du parcours scolaire ont été modifiées sur plusieurs aspects :

- mise à jour des modalités de la variable niveau de sortie (nomenclature diplôme) ;
- Limiter les oublis de libellé de diplômes sur certaines questions ;
- mention de tous les diplômes obtenus ;
- description fine du diplôme le plus élevé ;
- hiérarchisation explicite des diplômes.

#### b) La codification des diplômes et spécialités de formation

Traditionnellement, la codification des diplômes et des spécialités constituaient deux étapes distinctes. Pour la « Génération 2010 », le travail de codification a été réalisé simultanément.

Pour améliorer la codification des diplômes et des spécialités dans la « Génération 2010 », une combinaison de nouvelles nomenclatures issues de l'Insee a été utilisée :

- nomenclature des diplômes pour la codification de l'enquête emploi (version 01/01/2013) : détaillée en 104 postes (Annexe 3 p. 113) ;
- nomenclature des niveaux d'études utilisée pour la codification de l'enquête emploi (version 01/01/2013) : détaillée en 137 postes (Annexe 4 p. 117) ;
- nomenclature des groupes de Spécialités de Formation (NSF : niveau 100) (Annexe 5 p. 122).

Pour la première fois, une codification automatique à l'aide de Sicore a permis de réduire le temps consacré à la reprise manuelle, surtout pour la codification des diplômes (85 % codés

automatiquement). La mise en place de Sicore pour une codification dans la nomenclature Insee permet d'une part, d'améliorer la qualité de l'information et d'autre part, d'obtenir plus de cohérence en matière de calage de la pondération (calage sur les variables de l'enquête emploi).

En ce qui concerne le codage des spécialités de formation dans la nomenclature NSF, Sicore a été utilisé, hormis pour les sortants de rectorat ou de l'université pour lesquels l'information est connue dans la base de sondage.

### **Encadré 6 - SICORE Environnement diplôme et spécialité (millésime 2013)**

SICORE (Système Informatique de Codage des Réponses aux Enquêtes) est un système de chiffrage automatique développé par le Département des projets de l'Insee, qui a fait l'objet de tests sur de nombreuses variables et a déjà été appliqué en production avec succès. Son usage est appelé à se généraliser à l'Insee, et éventuellement à d'autres services statistiques, en France ou dans d'autres pays. A partir de l'enquête auprès de la « Génération 1998 », le Céreq a utilisé Sicore pour coder la profession. Avec la « Génération 2010 », son utilisation a été généralisée à la codification des autres grandes variables de l'enquête : activité, diplôme.

Sicore code les libellés de diplôme en utilisant deux variables annexes (formation secondaire/supérieure, année d'obtention du diplôme).

Le code renvoyé par Sicore est sur 8 positions :

- Positions 1 à 3 : renvoie au niveau du diplôme (ex : 320=BTS, 500=CAP, ...)
- Position 4 : zéro
- Positions 5 à 7 : spécialité selon le code NSF
- Position 8 : lettre

La nomenclature du plus haut diplôme du Céreq est la suivante. Elle tient compte du diplôme et de la spécialité.

01 : NON DIPLOME  
02I : CAP-BEP-MC INDUSTRIEL  
02T : CAP-BEP-MC TERTIAIRE  
03I : BAC PRO-BT-BP INDUSTRIEL  
03T : BAC PRO-BT-BP TERTIAIRE  
04I : BAC TECHNO INDUSTRIEL  
04T : BAC TECHNO TERTIAIRE  
05 : BAC GENERAL  
06I : BTS-DUT INDUSTRIEL  
06T : BTS-DUT TERTIAIRE  
07I : AUTRE BAC+2 INDUSTRIEL  
07T : AUTRE BAC+2 TERTIAIRE  
08 : BAC+2/3 SANTE SOCIAL  
09L : LICENCE PRO LSH GESTION DROIT  
09M : LICENCE PRO MATHS SCIENCE TECHNIQUE  
10L : L3 LSH GESTION DROIT  
10M : L3 MATHS SCIENCE TECHNIQUE  
11L : AUTRE BAC+3 LSH GESTION DROIT  
11M : AUTRE BAC+3 MATHS SCIENCE TECHNIQUE  
12L : BAC+4 LSH GESTION DROIT  
12M : BAC+4 MATHS SCIENCE TECHNIQUE  
13L : M2 LSH GESTION DROIT

13M : M2 MATHS SCIENCE TECHNIQUE  
14L : AUTRE BAC+5 LSH GESTION DROIT  
14M : AUTRE BAC+5 MATHS SCIENCE TECHNIQUE  
15 : BAC+5 ECOLE DE COMMERCE  
16 : INGENIEUR  
17 : DOCTORAT SANTE  
18L : DOCTORAT HORS SANTE LSH GESTION DROIT  
18M : DOCTORAT HORS SANTE MATHS SCIENCE TECHNIQUE

Une table de passage entre la nomenclature des diplômes de l'Insee, la nomenclature NSF et la nomenclature du Céreq a été créée.

### c) La codification du secteur d'activité de l'établissement employeur

Dans les enquêtes générations, le secteur d'activité des établissements était traditionnellement codé selon la nomenclature économique de synthèse (NES), dans sa version datant de 1994, désormais obsolète.

Avec la « Génération 2010 » à 3 ans, il a été décidé, en post-collecte, de coder désormais en NAF rev2 (nomenclature des activités économiques en vigueur en France depuis le 1<sup>er</sup> janvier 2008). Ce choix a été motivé par l'impossibilité de publier en 2014 dans une nomenclature obsolète depuis 2008. Désormais, l'utilisation de la NAF dans les enquêtes permettra, dans un souci de précision et de meilleure qualité de codification, la comparabilité de l'information avec les enquêtes françaises et européennes existantes. La NAF a la même structure que la nomenclature d'activités de la Communauté européenne (NACE révision. 2) mais elle comporte un niveau supplémentaire, spécifique à la France, celui des sous-classes. L'inconvénient majeur de ce choix est la perte de comparabilité avec les anciennes enquêtes Génération. Un travail de rétopolation sera programmé en 2015 si des besoins d'études et recherche sur le sujet se manifestent.

Cette codification a posé quelques difficultés car le questionnement n'était pas totalement adapté à une codification en NAF rev2. Pour la première fois, l'environnement Sicore activité a été mobilisé pour cette codification. La NAF révision 2 comporte cinq niveaux comprenant respectivement : 21, 88, 272, 615 et 732 postes. Avec les informations recueillies dans l'enquête, le choix s'est porté sur la nomenclature de niveau 2 (88 divisions).

Dans les futures enquêtes, une nouvelle série de questions appropriées pourra être intégrée avec la perspective d'une intégration de SICORE embarqué. Cet outil sera testé au moment de la ré-interrogation de la « Génération 2010 » à 5 ans et une expertise réalisée pour valider ce système (collecte / codification en direct) qui pourra être généralisée à l'ensemble des variables de l'enquête qui sont sujettes à des codifications post-enquête de grande ampleur.

#### La procédure « intermédiaire » mise en place pour « Génération 2010 » à 3 ans :

Dans les enquêtes générations, l'activité économique de l'établissement est définie à partir des questions EP2A à EP10. L'enchaînement des questions permet de définir l'activité *principale* et de recueillir l'information sur deux niveaux :

- un menu déroulant optimisé avec des noms d'entreprise/administration pour coder facilement l'activité en amont (sans question subsidiaire)
- à défaut une combinaison de questions permettant d'identifier l'activité en aval
  - soit par l'intermédiaire de questions fermées ;

- soit de questions fermées assorties d'une question ouverte contenant clairement la nature de l'activité.

Par conséquent, avec cet ensemble d'informations hétérogènes, trois méthodes de codage ont été utilisées séquentiellement :

- Imputation à partir d'une table de passage (30 % des périodes d'emploi) :

Pour réduire la passation du questionnaire et faciliter le travail à posteriori de codification, un menu déroulant a été construit au moment de la préparation de la « Génération 2004 » à 3 ans. Ce menu est une table contenant des noms de grandes entreprises du CAC 40 ainsi que des noms génériques d'administrations dont l'activité principale est facilement identifiée. Ainsi, lors de l'interrogation, si le nom de l'entreprise/administration cité existe dans ce menu, nous obtenions en théorie l'activité selon la nomenclature NES associée.

L'intérêt de ce menu est d'éviter des questions de description de l'entreprise (et donc du temps de questionnement) et d'éviter de poser des questions à l'enquêté dont la réponse est évidente dès lors que l'on a le nom de l'entreprise. En revanche, ce menu a plusieurs limites : les grands groupes détenant de nombreux établissements ayant des activités principales différentes, le menu était adapté à la NES et pas à la NAF rev2, certains sigles sont ambigus, le menu est un peu ancien même si celui-ci est toiletté à chaque enquête.

Avec la « Génération 2010 » à 3 ans, un travail a posteriori sur ce menu a consisté à coder ces noms d'entreprise/administrations selon la nomenclature NAF révision 2 (sur 88 divisions) et à imputer, en utilisant une table de passage, le code NAF correspondant sur l'ensemble des séquences d'emploi topées « information menu déroulant ».

- Imputation par codage automatique

Cas 1 : Traitement de l'information résultant d'un questionnement fermé (10 % des périodes d'emploi) :

Le questionnement, adapté à la logique de la construction de la NES et sous forme de questions fermées, permet de préciser la nature de l'activité et de trouver le code NES le plus approchant. Il s'agit de déterminer l'activité : est-elle liée à la fabrication ou au commerce de produits, alimentaires ou non alimentaires ? ou plutôt en lien avec de la vente de services ? pour des entreprises ou des particuliers ? A partir de ces informations, un codage automatique a été réalisé en interne. Si l'enquêté arrivait à se positionner dans cet enchaînement de questions (sans répondre « autre » à chaque fois), il ne renseignait pas l'activité en clair de son établissement.

Avec le changement de nomenclature décidé post enquête, le travail pour la « Génération 2010 » à 3 ans, a consisté à réfléchir à comment utiliser l'information recueillie pour déterminer le bon code NAF. L'outil Insee de recherche dans la nomenclature NAF s'est révélé précieux.

Il est disponible sur le site :

[http://recherche-naf.insee.fr/SIRENET\\_Template/Accueil/template\\_page\\_accueil.html](http://recherche-naf.insee.fr/SIRENET_Template/Accueil/template_page_accueil.html).

Si les questions fermées permettaient de conclure sur le code en NAF rev2, celui-ci était affecté. En revanche, si cela ne permettait pas de conclure, le nom de l'entreprise/administration et sa localisation étaient alors mobilisés pour trouver l'activité de l'établissement.

Cas 2 : Traitement de l'information résultant d'un questionnement fermé + information en clair de la nature de l'activité (30 % des périodes d'emploi) :

Dans le questionnement fermé, l'enquêté a la possibilité de préciser en clair la nature de l'activité dans une modalité « autre » lorsqu'il n'est pas parvenu à se positionner dans les modalités proposées.



La généralisation de la codification automatique avec l’outil SICORE dans l’enquête « Génération 2010 » (outil externe Insee) a facilité la transition pour aboutir à une codification de qualité selon la nomenclature NAF.

### Encadré 7 - SICORE Environnement Activité (millésime 2012)

Sicore permet de coder l’activité en NAF rév. 2. Cet environnement de codage est communément appelé Sicape (i.e. Sicore APE). Le codage n’utilise pas de variables annexes ; il s’appuie uniquement sur le libellé d’activité déclaré. Cet environnement Sicore a la particularité de retourner plusieurs codes APE, avec les probabilités associées. Pour cette raison, Sicape est utilisé sur le poste des enquêteurs dans les enquêtes Insee en face-à-face, car l’enquêté peut choisir le meilleur des codes proposés par Sicape. Mais en tenant compte des probabilités, on peut aussi utiliser Sicape en aval (traitement automatique). Dans ce cas, le taux de codage s’élève à 76 %. Mais le taux de bons codages est de 63 %.

Pour l’enquête Génération 2010, des règles de décisions ont été définies afin de repérer le meilleur écho.

- 1<sup>re</sup> règle : si Sicore fournit 1 SEUL écho dont la probabilité est supérieure à 40, l’écho en 5 positions est choisi.
- 2<sup>e</sup> règle : si Sicore fournit des échos multiples :
  - sélection des 2 premières positions des différents échos
  - calcul de la somme des probabilités des échos identiques sur 2 positions successives (premier écho inclus).

Dans ce deuxième cas, si la somme des probabilités est supérieure à 40, l’écho sur 2 positions est choisi. Le taux de bons codages est de 71,6 %.

#### - Codification manuelle

Pour les périodes d’emploi dont le secteur d’activité n’a pas pu être codé par les deux méthodes précédentes, un codage manuel :

- En interne (20 % des périodes d’emploi) : sur la base du nom de l’entreprise/administration et de sa localisation ;
- En externe (10 % des périodes d’emploi) : par une société extérieure à partir du nom de l’entreprise/administration, des informations en clair et des questions fermées (EP2A à EP10).

### d) La codification de la profession des périodes d’emploi

Dans l’enquête Génération, pour chaque séquence en entreprise longue, l’enquêté décrit sa profession à l’embauche et à la fin de la séquence (si celle-ci a évolué).

La profession a été codée selon la nomenclature des professions et catégories socioprofessionnelles (PCS) dans sa version actuelle datant de 2003.

Cette nomenclature a été créée par l’Insee pour disposer d’une classification pertinente pour l’analyse des comportements socio-économiques des personnes. Elle classe la population selon une synthèse de la profession (ou de l’ancienne profession), de la position hiérarchique et du statut (salarié ou non). La PCS comporte trois niveaux d’agrégation emboîtés : 8 groupes socioprofessionnels, 24 ou 42 catégories socioprofessionnelles et 486 professions.

La construction de la PCS s’appuie d’abord sur le libellé de l’emploi demandé à l’individu, complété par d’autres informations relatives à la position professionnelle de la personne et aux caractéristiques de l’entreprise dans laquelle elle travaille (appartenance au secteur public, taille de l’entreprise, secteur d’activité). Des corrections orthographiques et une normalisation ont été effectuées sur une partie des libellés, la plupart étant cependant issus d’un menu déroulant très détaillé proposé par l’application de collecte.

A partir de ces informations, le logiciel SICORE, permet de coder automatiquement la PCS. Pour l'enquête « Génération 2010 », les règles de décision ont été optimisées à partir de l'expertise d'une base de cas et de la notation de Sicore.

Toutes ces améliorations ainsi que l'intégration du secteur d'activité selon la nomenclature NAF rev2 (principale variable annexe) a permis d'atteindre un taux de codification de 82 % (supérieur à la moyenne antérieure de 70 %) et d'améliorer la qualité de l'information (le secteur d'activité NAF étant la principale variable intervenant dans les règles de décision).

Les professions qui n'ont pas pu être codées par cette procédure (18 % des séquences d'emploi) ont été transmises à une société externe qui a procédé à une codification manuelle.

### Encadré 8 - SICORE Environnement PCS 2013

Le logiciel Sicore permet d'effectuer un codage automatique des professions à partir d'un questionnaire en face-à-face ou bien à partir d'un fichier en entrée (fonctionnement en « Batch »). C'est le fonctionnement qui a été utilisé pour l'enquête « Génération 2010 ».

Sicore en Batch nécessite en entrée un fichier en format texte qui comporte les variables suivantes (cf. Annexe X : Définition des variables et des modalités)

- Identifiant
- Libellé de la profession
- STATUT : statut dans l'emploi (3 modalités)
- PUB : statut de l'établissement employeur (5 modalités)
- CPF : classification professionnelle ou qualification (10 modalités)
- FN : fonction principale (9 modalités)
- NBS : nombre de salariés employés (4 modalités)
- NAF4 : activité principale de l'établissement sur 4 caractères
- NAF2 : activité principale de l'établissement sur 2 caractères
- T : taille de l'entreprise (4 modalités)
- OPA : orientation des productions agricoles (8 modalités)
- DEP : département
- SAU : surface agricole utilisée (88 modalités)
- S : sexe
- SP : statut précaire
- STRE : emploi actuellement (Oui ou Non)

Il procède en 3 étapes successives :

- Etape 1 : il analyse le libellé de la profession
- Etape 2 : il entre dans une table de décision et fait éventuellement appel à certaines variables annexes (cf. la liste ci-dessus)
- Etape 3 : il donne en sortie au mieux un code PCS, ainsi qu'un indice de confiance du code proposé.

Dans le cadre de l'enquête « Génération 2010 », l'utilisation de la NAF2 a été privilégiée (la NAF4 n'étant que très partiellement présente) et les variables OPA, DEP, SAU et STRE non utilisées car informations non disponibles dans l'enquête. Toutes les autres variables ont été intégrées en variables annexes.

L'intégration de la NAF rev2 ainsi que l'application de nouvelles règles de décisions pour le choix du meilleur écho PCS (identiques à celles de la codification NAF) ont permis une augmentation du taux de bons codage (82%) ainsi qu'une amélioration de la qualité de l'information codée.

## e) La codification des contrats de travail

Pour les périodes d'emploi, plusieurs questions abordent la nature du contrat de travail en début et en fin de séquence, permettant par exemple de repérer les personnes qui accèdent à une entreprise via un emploi court (emploi aidé, mission d'intérim) puis y restent après passage en contrat à durée indéterminée.

Deux variables de synthèse sont donc construites pour définir le contrat de travail en début de séquence (stat\_emb) et le contrat de travail en fin de séquence (stat\_fin). Le tableau ci-dessous présente les différentes modalités existantes et les questions mobilisées pour les constituer.

**Tableau 25 – La codification du contrat de travail**

Modalité*	Nature du contrat	Questions mobilisées
01	Indépendant (à son compte)	Ep1
02	Aide familial	Ep1
03	Fonctionnaire	EP20 /EP26
04	CDI	EP20 /EP26
05	CDD sans distinction privé public	EP20 /EP26
06	Apprentissage	EP21 /EP27
07	Intérim	INTER, EP20 /EP26
08	Contrat de professionnalisation	EP21 /EP27
09	CAV	recodage de O_EP21/O_EP27
10	CAE, CEC	recodage de O_EP21/O_EP27
11	Autre contrat aidé (dont CIRMA, CIE)	recodage de O_EP21/O_EP27
12	Saisonnier	EP21 /EP27
13	Stagiaire	recodage de O_EP21/O_EP27
14	Pas de contrat	recodage de O_EP21/O_EP27
16	Autres	recodage de O_EP21/O_EP27
20	ATER	recodage de O_EP21/O_EP27
21	Post-doctorat + contrats recherche	recodage de O_EP21/O_EP27
22	Vacataire	recodage de O_EP21/O_EP27
23	Contrat étranger	recodage de O_EP21/O_EP27
24	Bourse	recodage de O_EP21/O_EP27
25	CNE	recodage de O_EP21/O_EP27
26	Intermittents du spectacle	recodage de O_EP21/O_EP27
27	CUI - CAE	EP21 /EP27
28	CUI - CIE	EP21 /EP27
29	Contrat d'avenir	EP21 /EP27
30	CDD secteur public	EP21 /EP27
31	CDD secteur privé	EP21 /EP27
32	Allocataire de recherche	recodage de O_EP21/O_EP27
33	Autre contrat aidé (CDDI)	recodage de O_EP21/O_EP27
34	Contrat engagement éducatif	recodage de O_EP21/O_EP27
35	Service civique	recodage de O_EP21/O_EP27
36	Volontariat	recodage de O_EP21/O_EP27

\*Ce découpage est le plus large existant. Pour les différentes publications, une agrégation de données a été utilisée.

## f) Les variables de localisation géographique

Pour l'ensemble des répondants, plusieurs localisations géographiques sont disponibles : la commune de résidence en 6<sup>e</sup>, au baccalauréat (le cas échéant), en 2010 et à la date de l'enquête en 2013, ainsi que le département de l'établissement de formation de l'année scolaire 2009-2010.

Pour chaque séquence en entreprise, la commune du lieu de travail est également renseignée.

Les communes sont repérées à partir de leur code officiel géographique, ce qui permet de mobiliser les différents zonages administratifs (région, département), et d'études développées par l'Insee (unités urbaines, aires urbaines, zones d'emploi). Les nomenclatures actualisées (en 2010) de l'Insee ont été intégrées pour cette nouvelle enquête Génération.

Pour le zonage en aires urbaines (ZAU), les informations retenues sont : la taille de l'aire urbaine et le type d'espace auquel la commune appartient (pôles urbain, couronne périurbaine, commune multipolarisée, communes isolées).

Chaque variable contenant une localisation géographique a été déclinée sous plusieurs variables, du format « nom de variable + nom du type code géographique ».

Voici un exemple de résultat pour la variable « sixième ».

- SIXIEMEDEP : Département de résidence de la commune en 6<sup>e</sup>
- SIXIEMEREG : Code de la région de résidence en 6<sup>e</sup>
- SIXIEMEREGIONF : Libellé de la région de résidence en 6<sup>e</sup>
- SIXIEMECATAEU : Catégorie d'espace du ZAU 2010 pour la commune de résidence en 6<sup>e</sup>
- SIXIEMETAU : Tranche d'aire urbaine 2010 pour la commune de résidence en 6<sup>e</sup>
- SIXIEMESTATUTUU : Statut de communes selon la définition des unités urbaines 2010 pour la commune de résidence en 6<sup>e</sup> (ville centre, banlieue, isolé, rural)
- SIXIEMETYPEUU : Type de commune selon la définition des unités urbaines 2010 pour la commune de résidence en 6<sup>e</sup> (rural/urbain)
- SIXIEMEZE : Code de la zone d'emploi de la commune de résidence en 6<sup>e</sup>
- SIXIEMELIBZE : Libellé de la zone d'emploi de la commune de résidence en 6<sup>e</sup>

## g) La codification des réponses en clair

Certaines variables présentent une modalité « Autre », qui ouvre la possibilité d'une réponse en clair. Dans la mesure du possible, les réponses en clair données par les individus ont été reclassées.

Une cinquantaine de variables ont fait l'objet de recodifications. Toute réponse ouverte donnée par un individu a été analysée. Lorsqu'un reclassement dans une modalité existante était possible, il a été effectué. Lorsque des modalités non proposées dans le questionnaire étaient citées de manière récurrente dans les ouverts, ces nouvelles modalités ont été créées. Les réponses n'ayant pas pu être reclassées selon les précédents critères sont restées dans la modalité « Autre ».

Les variables concernées sont signalées dans le dictionnaire des variables. Une indication « modalité ajoutée » est notée face aux nouvelles modalités créées.

La liste des variables ouvertes est disponible en annexe 9.

## 4. Le traitement des salaires (primes incluses)

Au moment de l'interrogation, on demande systématiquement aux individus, pour chaque séquence d'activité déclarée, le salaire net perçu à l'embauche et le salaire de fin de séquence ou actuel (si augmentation du salaire dans la séquence). Uniquement les observations des salariés sont concernées par cette phase d'imputation. Les séquences d'activité correspondant aux situations d'indépendants ou du statut d'aide familiale ne sont pas prises en compte puisqu'il s'agit de revenus. Pour ces derniers,

les revenus fournis dans les bases sont REVDEB et REVFIN (information brute). Cette question sur les salaires est dite "sensible", on donne donc la possibilité aux enquêtés de répondre soit en clair, soit par tranches de salaire. Ils ont aussi la possibilité de ne pas répondre. Les structures du type de déclaration des salaires sont les suivantes pour les débuts et les fins des séquences d'emploi.

**Structures du type de déclaration pour les débuts de séquence**

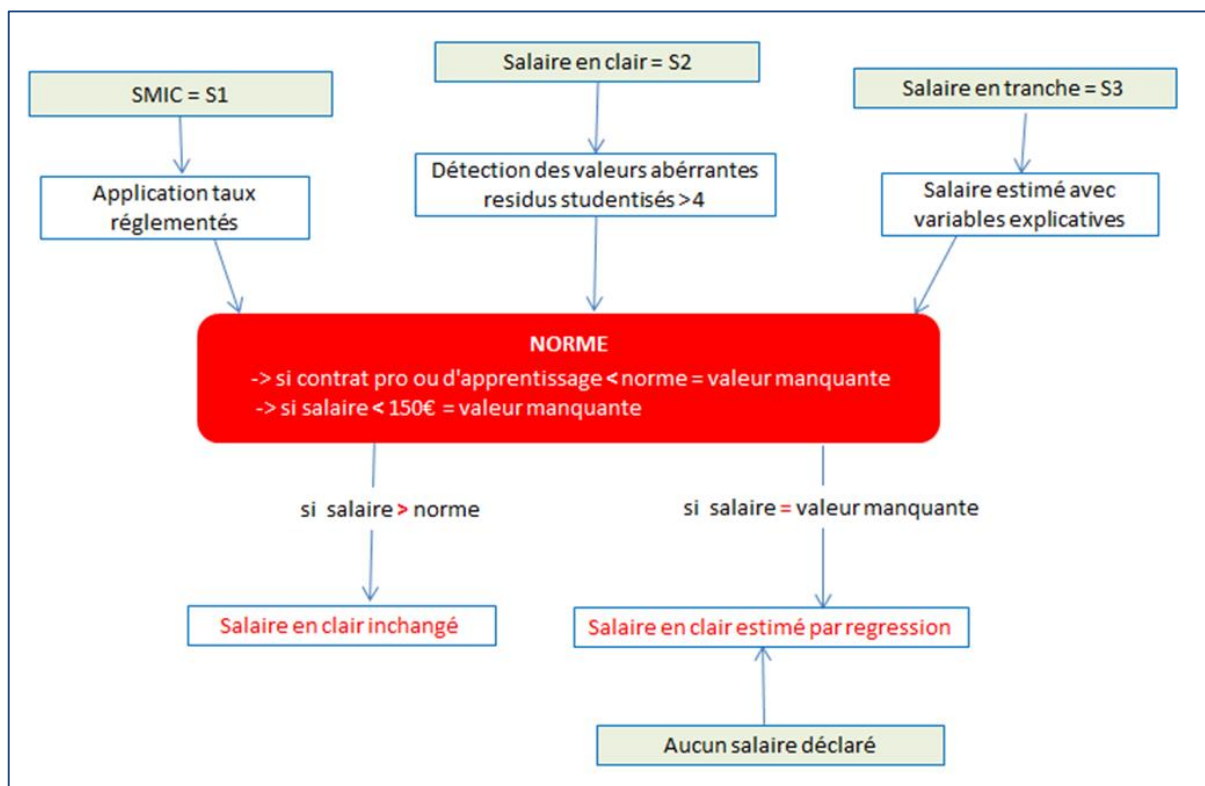
Type de déclaration	Nombre de séquences	Pourcentage
Clair	44 922	77,69
Smic	7 368	12,74
Tranche	5 188	8,97
Manquant	343	0,59

**Structures du type de déclaration pour les fins de séquence**

Type de déclaration	Nombre de séquences	Pourcentage
Clair	45 545	78,77
Smic	6 864	11,87
Tranche	4 628	8,00
Manquant	784	1,36

Le traitement des salaires se fait en fonction du mode de réponse. Le déroulement des étapes est présenté dans le schéma suivant :

**Figure 16 - Gestion des salaires**



Tous les salaires sont rapportés à une NORME construite à partir de la valeur du SMIC d'une période donnée (5 revalorisations du SMIC sur les trois années observées) mais aussi en fonction de la quotité de travail. Pour les contrats pro et apprentissage, on intègre l'âge et l'avancement dans la formation.

**Procédure de calcul**

Pour l'ensemble des modèles estimés, on a utilisé l'option STEPWISE (sélection du meilleur modèle à partir de critères d'informations). Contrairement à « Génération 2004 », deux modèles distincts sur les salaires manquants (temps plein/temps partiel) ont été spécifiés. Par ailleurs, pour les salaires déclarés en tranche, on a estimé les salaires à partir d'un modèle alors qu'auparavant on utilisait le centre des classes.

**Tableau 26 - Définition du modèle type sur les salaires (avec évolution inter-enquête)**

« Génération 2004 » à 3 et 5 ans (7 variables)	« Génération 2004 » à 7 ans (10 variables)	« Génération 2010 » à 3 ans (14 variables)
l'âge de l'enquêté (age04)	l'âge de l'enquêté (age04)	l'âge de l'enquêté (age10)
le statut de l'entreprise public ou privé (ep3)	le statut de l'entreprise PUB est remplacé par la variable NES en 8 postes	le statut de l'entreprise codé en NES est remplacé par la variable NAF en 8 postes
la taille pour les petites entreprises de moins de 20 salariés (ep12)	la taille pour les petites entreprises de moins de 20 salariés (ep12)	la taille pour les petites entreprises de moins de 20 salariés (ep12)
la catégorie socio-professionnelle de l'enquêté (pospro) : les ouvriers, les professions intermédiaires, les cadres, les employés, autres	la catégorie socio-professionnelle de l'enquêté (pospro) : les ouvriers, les professions intermédiaires, les cadres, les employés, autres	la catégorie socio-professionnelle de l'enquêté (pospro) : les ouvriers, les professions intermédiaires, les cadres, les employés, autres
le temps de travail : temps plein, 80 %, 60 %, 50 %, 40 %, 20 %, moins de 8h, indéterminé	le temps de travail : temps plein, 80 %, 60 %, 50 %, 40 %, 20 %, moins de 8h, indéterminé	2 modèles utilisés : - Temps plein - Temps partiel 80 %, 60 %, 50 %, 40 %, 20 %, moins de 8h
le contrat de travail : contrats aidés, cdi, indépendant/ aide familiale, contrats type fonctionnaire, tous contrats type cdd	le contrat de travail : contrats aidés, cdi, indépendant/ aide familiale, contrats type fonctionnaire, tous contrats type cdd	le contrat de travail : contrats aidés, cdi, indépendant/ aide familiale, contrats type fonctionnaire, tous contrats type cdd
le sexe (q1)	le sexe (q1)	le sexe (q1)
	l'expérience déterminée par le nombre de mois travaillés avant l'entrée dans l'emploi (moisx)	l'expérience (moisx) <i>Non sélectionnée avec l'option stepwise</i>
	le niveau de sortie de l'enquêté (nivsor9 en 5 catégories) : Non diplômés, secondaire, bac+2, bac+3/4, bac+5	le niveau de sortie de l'enquêté (nisor en 5 catégories) : Non diplômés, secondaire, bac+2, bac+3/4, bac+5
	la région de l'entreprise (reg) : ldf, autres régions, étranger	la région de l'entreprise (reg): ldf, autres régions, étranger
		plus haut diplôme obtenu (phd en15 positions)
		qualification (sanq : indicatrice valant « 1 » si le sortant n'a pas de qualification)
		spécialité de formation (NSF en 3 catégories) : général, industriel et tertiaire
		Ancienneté (durée)
		origine des parents (lieunmer, lieunper) <i>Non sélectionnée avec l'option stepwise</i>

En rouge : les nouvelles variables intégrées au modèle par rapport à l'enquête précédente.

En vert : les variables testées et non sélectionnées par l'option STEPWISE.

**Etape 1** : détermination des salaires aberrants

Une première détermination des salaires aberrants a été faite à partir d'un modèle de régression basé sur les variables sexe, âge, niveau de sortie, indicatrice qualification, plus haut diplôme, spécialité, type de contrat, catégorie sociale, région, taille de l'entreprise, activité de l'entreprise, ancienneté. A l'issue de ce modèle, les observations dont les résidus studentisés sont supérieurs à 4, ont été considérées comme aberrantes, et les salaires correspondants ont été mis à valeur manquante. Cette opération concerne environ 300 séquences que ce soit pour les salaires de début ou de fin.

**Etape 2** : Imputation des salaires déclarés à partir des tranches

Lorsque le salaire a été déclaré via les tranches, un modèle de régression mobilisant les variables suivantes a été spécifié : sexe, âge, niveau de sortie, une indicatrice portant sur le fait de ne pas avoir de qualification, le plus haut diplôme, le type de contrat, la catégorie sociale, l'activité de l'entreprise, la région, l'ancienneté, « la classe de la tranche de salaire ». Ce modèle a été estimé sur les observations ayant le salaire déclaré en clair. La variable à expliquer de ce modèle est le salaire en clair, les tranches de salaires ont été recalculées à partir des salaires en clair.

**Etape 3** : Comparaison des salaires déclarés aux normes en vigueur.

Les salaires déclarés sont comparés à une norme « basse » en fonction de la quotité de travail et du statut et calculés à partir du taux horaire du SMIC net et des taux réglementés (pour les contrats spécifiques tels que le contrat de professionnalisation et le contrat d'apprentissage). Si les salaires sont inférieurs à cette norme, ils sont mis à valeur manquante, dans le cas contraire ils sont conservés.

Le SMIC a été réévalué 5 fois entre juin 2010 et juillet 2013 (période d'interrogation)

**Tableau 27 – Revalorisation du SMIC entre juin 2010 et juillet 2013**

Date de revalorisation	Montant net du SMIC
Janvier 2010	1 056,24
Janvier 2011	1 072,07
Décembre 2011	1 094,71
Janvier 2012	1 096,88
Juillet 2012	1 118,29
Janvier 2013	1 120,43

\*Source INSEE

**Le contrat de professionnalisation** (nouveau contrat de qualification)

Ce contrat s'adresse aux jeunes de 16 à 25 ans, aux demandeurs d'emplois de 26 ans et plus et aux bénéficiaires de certaines allocations ou contrats. Les bénéficiaires de 16 à 25 ans révolus sont rémunérés en pourcentage du smic selon leur âge et leur niveau de formation. Les salariés âgés de 26 ans et plus perçoivent une rémunération qui ne peut être, ni inférieure au smic, ni à 85 % du salaire minimum conventionnel.

Le calcul du salaire prend en compte l'âge et le niveau de formation pour chaque individu. La variable nisor (niveau de sortie sur 15 positions) est utilisée comme référence pour le niveau de formation.



**Tableau 28 – Normes de rémunérations des contrats de professionnalisation**

	Inférieur au BAC	BAC et plus
<b>Moins de 21 ans</b>	55 %	65 %
<b>21 ans et plus</b>	70 %	80 %

\*pourcentage du smic

Ce salaire ne peut être inférieur à 55 % du smic pour les bénéficiaires âgés de moins de 21 ans et à 70 % du smic pour les bénéficiaires de 21 ans et plus. Ces rémunérations ne peuvent être inférieures, respectivement à 65 % et 80 % du smic, dès lors que le bénéficiaire est titulaire d'une qualification au moins égale à celle d'un baccalauréat professionnel ou d'un titre ou diplôme à finalité professionnelle de même niveau.

#### Le contrat d'apprentissage

Ce contrat s'adresse aux jeunes de 16 à 25 ans mais des dérogations à ces limites d'âges sont possibles. L'apprenti perçoit un salaire déterminé en pourcentage du smic et dont le montant varie en fonction de l'âge du bénéficiaire et de sa progression dans le ou les cycles de formation, faisant l'objet de l'apprentissage individuel.

**Tableau 29 - Normes de rémunérations des contrats d'apprentissage**

Année d'exécution du contrat	Moins de 18 ans	De 18 ans à moins de 21 ans	21 ans et plus
<b>1<sup>re</sup> année</b>	25 %	41 %	53 %
<b>2<sup>e</sup> année</b>	37 %	49 %	61 %
<b>3<sup>e</sup> année</b>	53 %	65 %	78 %

\*pourcentage du smic

Au titre de sa progression dans un cycle de formation, l'apprenti bénéficie d'une rémunération variant en fonction de l'année d'exécution du contrat. Ainsi, le salaire minimum perçu par l'apprenti correspond au pourcentage du smic (ou, dans certains cas, du salaire minimum conventionnel), allant de 25 % à 78 % déterminé en fonction de son âge et de sa progression dans le cycle.

Pour déterminer l'année d'exécution du contrat, on suppose que la période d'emploi sur laquelle il déclare être apprenti correspond à son lieu d'apprentissage. La variable durée est utilisée par défaut comme paramètre pour déterminer l'année d'exécution du contrat.

Un test de cohérence est réalisé comme pour les autres contrats pour vérifier si le salaire déclaré est proche du salaire calculé en fonction des conditions de rémunération dans le cadre d'un contrat d'apprentissage ou d'un contrat de professionnalisation.

#### **Etape 4** : Imputation des salaires manquants

Les salaires inférieurs à 150 € et les salaires manquants sont imputés par un modèle faisant intervenir les variables : sexe, âge, niveau de sortie, indicatrice qualification, plus haut diplôme, spécialité, type de contrat, catégorie sociale, région, taille de l'entreprise, activité de l'entreprise, ancienneté modèle séparé pour le temps partiel et les temps complets. Les modèles de régression expliquent environ la moitié de la variance totale. Ces imputations ont été faites en séparant les observations à temps plein des observations à temps partiel.

#### **Etape 5** : Le traitement des primes

Comme pour les salaires, des questions sur les primes sont posées à chaque interrogation aux individus qui déclarent un salaire hors primes. Les enquêtés ont la possibilité de nous déclarer des primes en clair, en tranches, ou refuser de nous déclarer leur montant.

*Remarque : dans les bases mises à disposition, les informations sur les primes ne sont pas fournies car intégrées directement dans le calcul du salaire d'embauche ou de fin.*

Il existe deux types de primes :

- La prime dite de « 13<sup>e</sup> mois » : si l'individu déclare toucher un 13<sup>e</sup> mois, on calcule une prime mensuelle en fonction du salaire déclaré corrigé. On ajoute cette prime au salaire.
- Les « autres primes » : on collecte des primes mensuelles en clair ou en tranches. On applique le même traitement que pour le salaire. On détermine une moyenne par tranche et par statut obtenue à partir des autres primes déclarées en clair et on ajoute au salaire une prime moyenne en clair.

L'objectif étant de déterminer un salaire « primes incluses » en clair pour toutes les séquences d'emplois (hormis les jobs de vacances et les emplois de statut aide familiale).

#### **Etape 6** : le traitement des salaires des extensions

Le traitement des salaires des extensions suit une logique similaire. Les modèles ont été réalisés sur les sortants appartenant au champ spécifique des extensions uniquement.

## 5. Les tables finales

A l'issue des différents traitements, les fichiers résultats de l'enquête « Génération 2010 » à trois ans se répartissent en trois tables.

Ces tables concernent les personnes du champ Céreq qui ont répondu au questionnaire. Les hors champ Céreq satisfaisant à des besoins d'extension (post-initiaux) ont été placés dans des tables séparées.

Nom	L'observation dans la table est :	Contenu
<b>Table « Individu »</b>	Le répondant (33 547 observations)	Ensemble des réponses de l'enquêté, y compris le calendrier professionnel et le mode d'habitation. Ne sont pas décrites les parties descriptives des séquences d'emploi et de non emploi. Informations issues de la base de sondage. Pondérations. Environ 700 variables.
<b>Table « Séquences d'emploi »</b>	La période d'emploi dans l'entreprise (62 149 observations)	Ensemble des périodes d'emploi décrites dans le calendrier professionnel et variables descriptives de ces périodes. Un identifiant (IDENT) permet d'associer les périodes au répondant concerné de la table « Individu ». Un numéro de séquence (NSEQ) permet de distinguer les séquences de chaque individu et de connaître la chronologie de ses séquences. Environ 100 variables.
<b>Table « Séquences de non emploi »</b>	La période de non emploi (29 747 observations)	Ensemble des périodes de non emploi, décrites dans le calendrier professionnel et variables descriptives de ces périodes. Un identifiant (IDENT) permet d'associer les périodes au répondant concerné de la table « Individu ». Un numéro de séquence (NSEQ) permet de distinguer les séquences de chaque individu et de connaître la chronologie de ses séquences. Environ 100 variables.

Trois extensions de champ intégrant les post-initiaux et levant le seuil d'âge de 35 ans ont été réalisées lors de l'enquête « Générations 2010 » à trois ans. L'une, relative aux formations sanitaires et sociales, pour la Drees ; l'autre relative aux formations du sport, pour le ministère de la Jeunesse, des sports et de la vie associative, et pour finir, la troisième relative aux formations des agents de la fonction publique pour la Dgafp.

Chacune de ces trois extensions a donné lieu à la production des trois tables « Individus », « Séquences d'emploi » et « Séquences de non emploi », sur l'ensemble des individus dans leurs champs respectifs, qu'ils soient primo sortants ou post-initiaux. Chacune dispose également d'une variable de pondération spécifique.

Les tables contiennent les variables spécifiques au module « post-initiaux » et, pour l'extension des formations du sport, les réponses à l'extension de questionnement.

### **Les bases comparables**

Pour usage interne, le Céreq a réalisé une base nationale à champ comparable pour les enquêtes à 3 ans des Générations 2004, 2007, 2010.

Les principaux traitements réalisés sont les suivants :

- Définition d'une nomenclature de diplôme agrégée commune aux trois générations
- Recalcul des pondérations sur les générations précédentes (lié à l'affinement des marges de calage)
- Suppression des Dom pour la « Génération 2010 »
- Suppression des sortants d'IUFM pour les Générations 2004 et 2007.

Il est important de noter que la comparaison d'une même variable nécessite tout de même de revenir aux formulations et aux modalités de la question dans chacune des enquêtes.

## VII. La pondération finale

### 1. Le principe général

Pour « Génération 2010 », un total de 33 547 questionnaires a été collecté auprès d'individus appartenant au champ du Céreq. Pour ces individus, la pondération finale est obtenue en tenant compte des paramètres suivants :

En amont de la collecte :

- couverture de la base de sondage ;
- probabilité individuelle d'appartenance à l'échantillon principal.

En aval de la collecte :

- probabilité d'entrer en contact téléphonique avec l'individu ou un de ses proches sachant que l'individu appartient à l'échantillon principal ;
- probabilité que l'individu accepte de répondre, sachant que lui-même ou un proche a été contacté ;
- probabilité de terminer le questionnaire sachant que l'individu a accepté de répondre au questionnaire ;
- coefficient de calage sur marges.

La correction du taux de couverture vise à pallier les défauts d'exhaustivité de la base de sondage, qui résultent de la non-réponse de certains établissements à la phase de collecte des fichiers d'inscrits.

La probabilité d'appartenance à l'échantillon est égale, pour les individus de l'échantillon utilisé, à leur probabilité d'être sélectionnés dans la phase de tirage de l'échantillon. Pour « Génération 2010 », l'échantillon de réserve n'a pas été mobilisé et seul l'échantillon principal a été utilisé. La probabilité d'appartenance à l'échantillon est donc égale à la probabilité d'appartenance à l'échantillon principal. Pour déterminer la probabilité de répondre tout en étant dans le champ et en ayant terminé le questionnaire, trois modèles successifs sont mobilisés.

La probabilité d'entrer en contact téléphonique avec l'individu ou un de ses proches a été calculée pour l'échantillon principal<sup>10</sup>.

La probabilité d'accepter de répondre est calculée pour les individus de l'échantillon principal avec lesquels le contact a pu être établi et qui ne sont pas connus comme résidant à l'étranger.

La probabilité de terminer le questionnaire est calculée pour les individus ayant accepté de répondre et qui appartiennent au champ du Céreq.

Un coefficient de calage sur marge a été calculé. Pour le champ Céreq, le calage est fait à partir de résultats publiés par la DEPP<sup>11</sup> qui mobilisent l'enquête emploi de l'Insee.

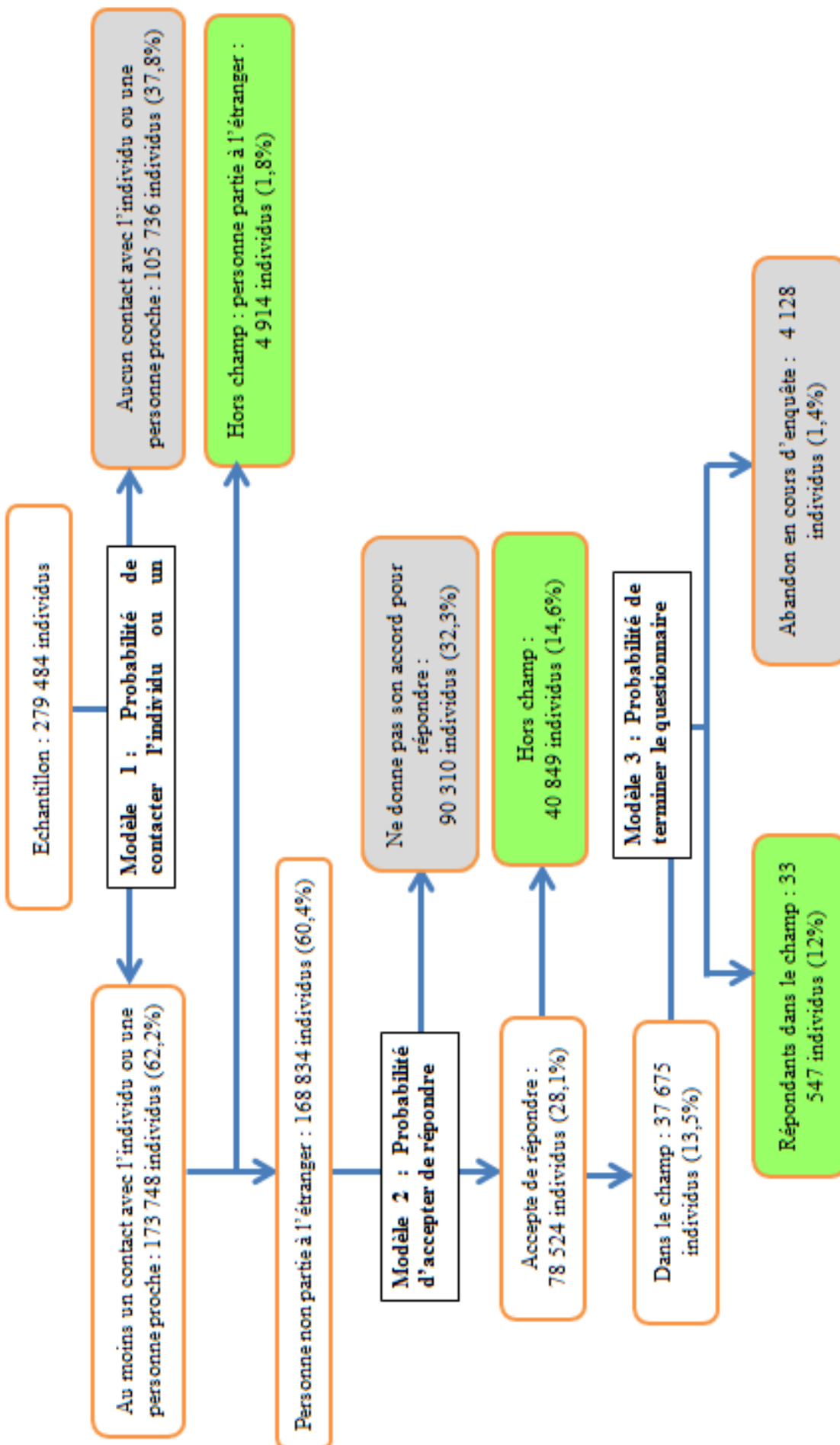
Finalement, le poids d'un individu répondant dans le champ du Céreq est déterminé de la manière suivante :

$$\forall i, \text{pond}ef(i) = p_{\text{couverture}}(i) * p_{\text{échantillon}}(i) * p_{\text{contact}}(i) * p_{\text{accepte répondre}}(i) * p_{\text{termine questionnaire}}(i) * p_{\text{calage}}(i)$$

<sup>10</sup> L'introduction de cette étape de modélisation a été effectuée notamment pour permettre de détecter les individus partis à l'étranger qui ne font pas partie du champ de l'enquête CATI mais qui ont cependant été interrogés par ailleurs via un questionnaire internet.

<sup>11</sup> Ouvrage « Repères et références statistiques sur les enseignements, la formation et la recherche » – RERS 2013 (fiche Le niveau d'étude à la sortie du système éducatif).

Figure 12. Schéma général de la modélisation de la non-réponse

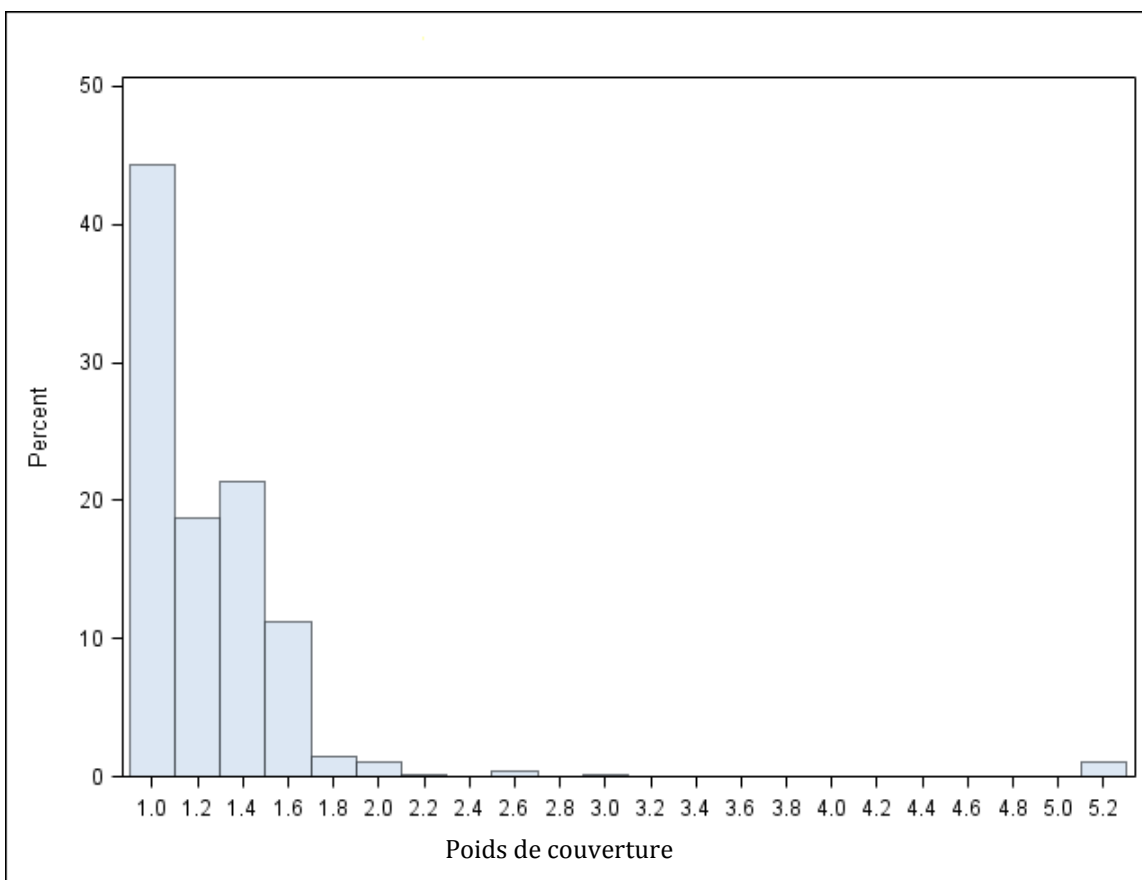


## 2. Le poids de couverture

Le calcul du poids de couverture est présenté dans la partie II2.0 (page 25). Ce calcul est réalisé sur l’ensemble des sortants avec l’hypothèse sous-jacente d’une homogénéité des taux de couverture entre les individus dans le champ Céreq et les individus hors champ.

Pour les 33 547 individus dans le champ de l’enquête et répondant au questionnaire, il est compris entre 1 (dans le cas où la base de sondage est considérée comme exhaustive) et 5,2<sup>12</sup>.

**Figure 17 Distribution des poids de correction liés au défaut de couverture de la base de sondage**

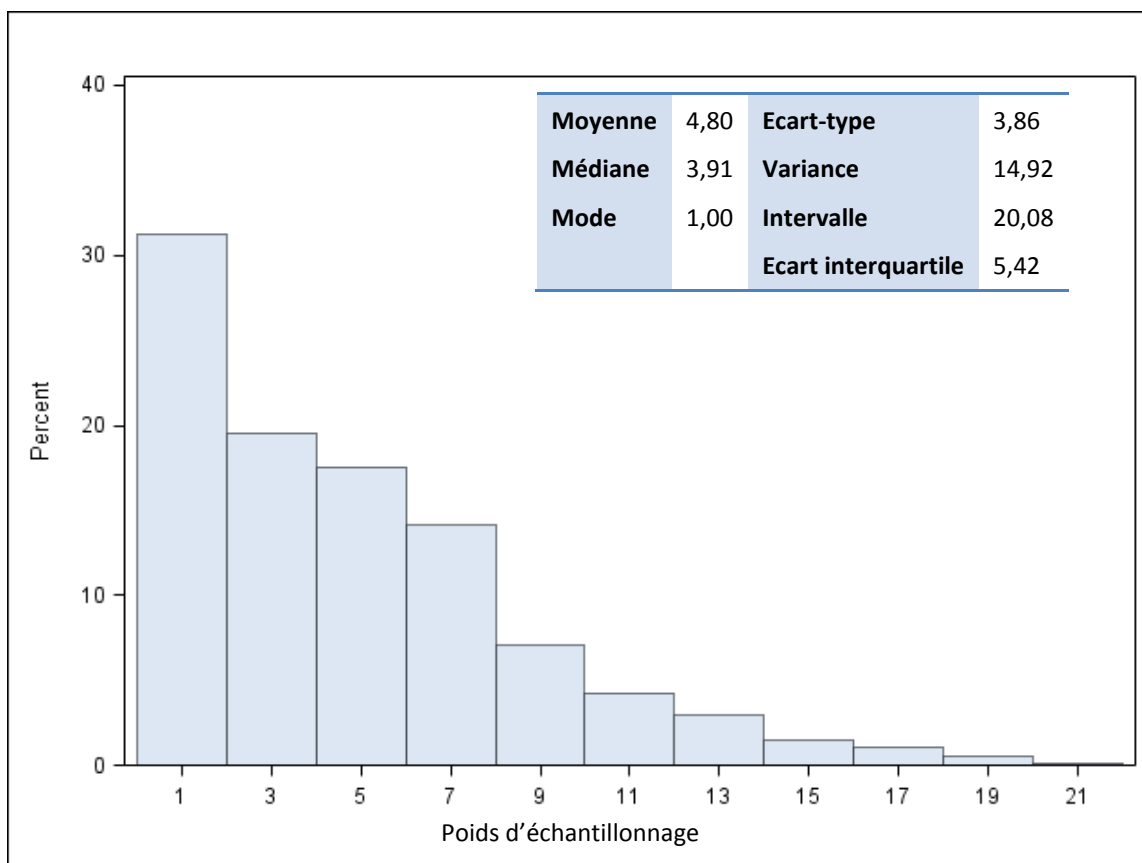


<sup>12</sup> Ces poids de couverture atypique concernent quelques spécialités du champ santé/social (pédicure-podologue, ...)

### 3. Le poids d'échantillonnage

Pour les individus dans le champ de l'enquête et répondant au questionnaire, il est compris entre 1 (l'individu est échantillonné d'office) et 21,1.

**Figure 18 Distribution des poids d'échantillonnage**



## 4. Le poids relatif au fait de contacter l'individu ou un proche

### a) Modélisation de la probabilité de contact

L'échantillon est constitué de 279 484 individus. Un contact téléphonique a pu être établi pour 173 748 d'entre eux (62,2 %). La probabilité de contact n'est pas indépendante des caractéristiques de l'individu connues dans l'échantillon. On modélise la probabilité de contact grâce à une régression logistique sur l'échantillon. Les variables explicatives suivantes sont introduites dans le modèle :

- strate de formation
- âge de l'individu (en 2010) (non retenu par le stepwise)
- variables indicatrices sur le type de coordonnées téléphoniques disponibles dans l'échantillon (numéro de téléphone fixe, portable, issu de la base de sondage ou de recherches de coordonnées téléphoniques, ...)
- mode d'envoi de la lettre-avis
- présence d'un mail dans la base de sondage (non retenu dans le stepwise mais lié au mode d'envoi de la lettre-avis)
- déménagement entre 2010 et 2013
- formation effectuée par apprentissage en 2010
- caractéristiques de la commune de résidence en 2010 (ZAU ou étranger)
- présence du nom de la commune sur l'adresse de l'individu en 2010
- présence d'un code postal sur l'adresse de l'individu en 2010
- département de l'établissement de formation en 2010
- information sur l'adresse de l'individu en 2010 (zus, / non zus ou inconnu)
- genre

Les individus dont les coordonnées téléphoniques disponibles étaient les plus fiables ont été les plus faciles à contacter. Ceux dont l'adresse postale en 2010 venait d'un pays étranger, était située en ZUS ou était mal renseignée (code postal absent, ...) ont été moins souvent contactés que les autres. Les jeunes résidant dans une commune urbaine en 2010 ont également moins souvent fait l'objet d'un contact. Le département de formation et le type de formation suivie sont également liés à la probabilité de joindre l'individu ou un de ses proches.

### b) Cohérence de la modélisation

Le modèle présente un  $R^2$  ajusté de 0,29 et 77 % de paires concordantes.

Sur l'ensemble de l'échantillon, la probabilité moyenne de contact estimée est égale à 0,62 (min=0,03 ; max=0,99 ; ET=0,23).

Pour les 173 748 individus ayant effectivement été contactés, le poids relatif au contact est égal à l'inverse de la probabilité de contact estimée par la procédure logistique.

La somme des poids relatifs au contact pour les individus effectivement contactés est égale à 272 540. On note une légère contraction des poids par rapport à l'échantillon (279 484 individus, soit un écart de 2 %).

Grâce à la modélisation mise en œuvre, la pondération obtenue sur les individus contactés permet de limiter les déformations (que ce soit au niveau de l'échantillon, de la base de sondage, ou de la base de sondage redressée de la couverture) sur les variables telles que le genre, le fait d'habiter en ZUS, la strate de formation, la région de l'établissement de formation.

Par contre, malgré la présence de la variable caractéristique de la commune de résidence dans le modèle, la modélisation n'est pas très performante sur les individus qui avaient une adresse étrangère ou inconnue en 2010 (le nombre d'individus possédant une adresse à l'étranger est estimé à 10 990 sur la base de sondage corrigée de la couverture et à 9 610 après correction du contact (-12 %) ; le nombre d'individus possédant une adresse inconnue est estimé à 30 630 sur la base de sondage corrigée de la couverture et à 26 710 après



correction du contact (-13 %). Ce constat n'est pas surprenant car cette population est par construction difficile à atteindre par téléphone et sans doute fréquemment hors du champ de l'enquête (étudiants étrangers rentrés dans leur pays). C'est probablement à ce niveau que l'on observe la contraction des poids.

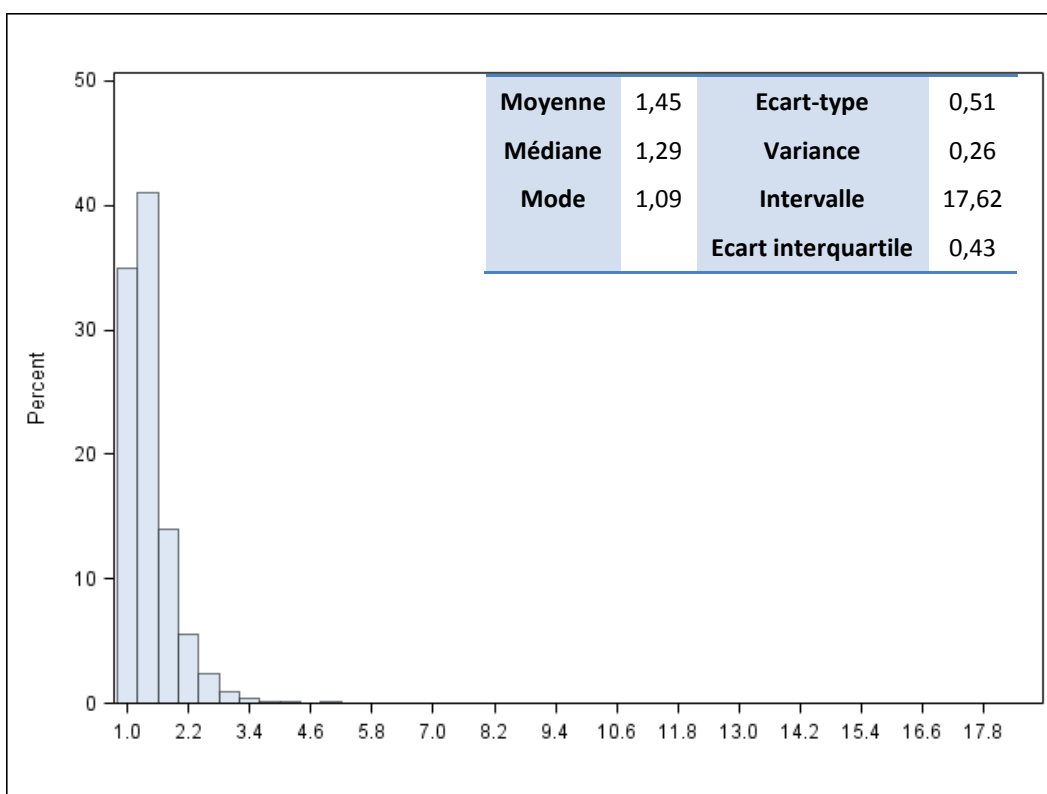
### c) Mise hors champ des individus qui résident à l'étranger

Parmi les 173 748 individus pour lesquels un contact téléphonique a pu être établi avec le jeune ou une personne proche, 4 914 individus se sont révélés résider à l'étranger et donc être hors du champ de l'enquête. Ces individus sont donc considérés comme répondants hors champ et exclus de la suite de l'analyse. En données pondérées, ils représentent 40 000 individus sur la base de sondage corrigée du défaut de couverture.

### d) Poids de contact des individus ayant complété un questionnaire dans le champ du Céreq

Pour les 33 547 individus dans le champ de l'enquête et répondant au questionnaire, il est compris entre 1 et 18,6.

**Figure 19 - Distribution des poids de contact**



## 5. Le poids relatif au fait d'accepter de répondre

### a) Modélisation de la probabilité d'accepter de répondre

A l'issue de la phase de contact et après exclusion des individus partis à l'étranger (hors du champ), il reste 168 834 individus potentiellement dans le champ du Céreq.

Parmi eux, 78 524 (46,5 %) ont accepté de répondre au questionnaire et ont complété la partie filtre. On modélise à nouveau la probabilité d'accepter de répondre en fonction des caractéristiques individuelles grâce à une régression logistique. Les variables explicatives suivantes qui sont introduites dans le modèle sont les mêmes que celles introduites dans le modèle de contact, à savoir :

- strate de formation
- âge de l'individu (en 2010)
- variables indicatrices sur le type de coordonnées téléphoniques disponibles dans l'échantillon (numéro de téléphone fixe, portable, issu de la base de sondage ou de recherches de coordonnées téléphoniques, ...)
- mode d'envoi de la lettre-avis
- présence d'un mail dans la base de sondage (pas retenu stepwise)
- déménagement entre 2010 et 2013 (pas retenu stepwise)
- formation effectuée par apprentissage en 2010
- caractéristiques de la commune de résidence en 2010
- présence du nom de la commune sur l'adresse de l'individu en 2010 (pas retenu stepwise)
- présence d'un code postal sur l'adresse de l'individu en 2010 (pas retenu stepwise)
- département de l'établissement de formation en 2010
- information sur l'adresse de l'individu en 2010 (zus, non zus, inconnu) (pas retenu stepwise)
- genre (pas retenu stepwise)

Les individus dont les coordonnées téléphoniques disponibles étaient les plus fiables ont plus souvent acceptés de répondre (pour les autres, on peut supposer que le premier contact téléphonique a plus fréquemment eu lieu auprès d'une tierce personne, ne permettant pas d'atteindre l'individu *in fine*). Les individus plus âgés et ceux qui ont suivi une formation par apprentissage acceptent moins fréquemment de répondre. La strate de formation et le département de l'établissement de formation ont également leur importance.

### b) Cohérence de la modélisation

Le modèle présente 60 % de paires concordantes. Le pouvoir prédictif du modèle est donc faible, nettement inférieur à celui du modèle de contact précédent. Si les caractéristiques individuelles jouent un rôle majeur sur la probabilité de joindre un individu ou un proche, le fait d'accepter de répondre sachant qu'on a été contacté, en dépend moins.

Sur l'ensemble des 168 834 individus ayant fait l'objet d'un contact (hors individus partis à l'étranger), la probabilité moyenne d'acceptation estimée est égale à 0,47 (min=0,14 ; max=0,81 ; ET=0,09).

Pour les 78 524 individus ayant effectivement accepté de répondre, le poids relatif est égal à l'inverse de la probabilité d'accepter de répondre estimée par la procédure logistique.

### c) Mise hors champ des individus à l'issue du questionnaire filtre

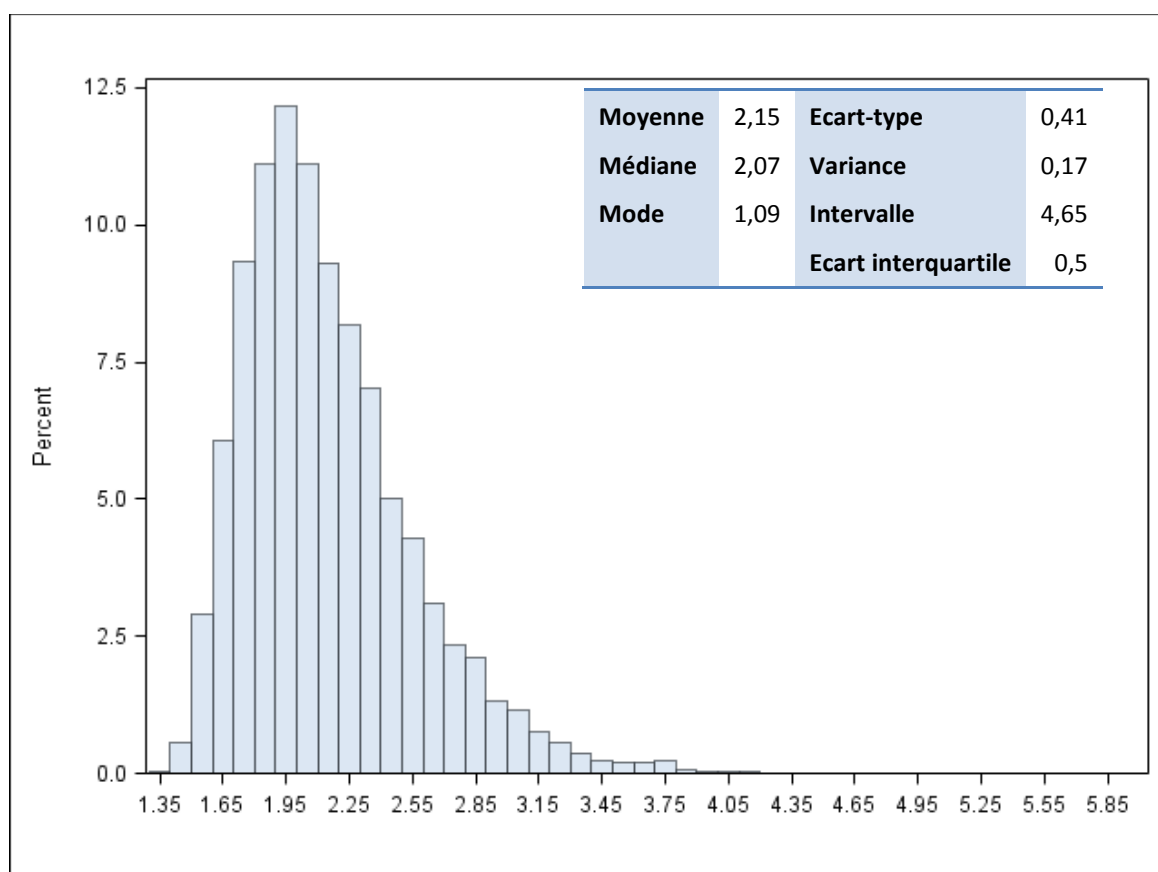
Un total de 40 849 individus ayant accepté de répondre s'est révélé hors du champ de l'enquête. Il s'agit principalement d'individus ayant poursuivi leurs études à la rentrée scolaire 2010-2011, ou d'individus post-initiaux (ayant interrompu leurs études une première fois puis les ayant reprises avant de les arrêter à nouveau en 2009-2010). Ces individus sont donc considérés comme répondants hors champ et exclus de la

suite de l'analyse. En données pondérées, ils représentent 632 500 individus sur la base de sondage corrigée du défaut de couverture (environ 1 400 000).

#### d) Poids relatif au fait d'accepter de répondre pour les individus ayant complété un questionnaire dans le champ du Céreq

Pour les 33 547 individus dans le champ de l'enquête et répondant au questionnaire, il est compris entre 1,31 et 5,97.

**Figure 20 - Distribution des poids relatifs au fait d'accepter de répondre**



## 6. Probabilité de répondre à l'intégralité du questionnaire sachant que l'on appartient au champ de l'enquête

### a) Modélisation de la probabilité de terminer le questionnaire

A l'issue du questionnaire filtre et après exclusion des individus hors du champ, 37 675 personnes ont commencé un questionnaire.

Parmi elles, 33 859 (90 %) ont terminé le questionnaire. On modélise la probabilité de terminer un questionnaire sachant que l'on est dans le champ du Céreq, en fonction des caractéristiques individuelles grâce à une régression logistique. Les variables explicatives introduites dans le modèle sont les mêmes que celles des modèles précédents. Les variables retenues sont la strate de formation, le fait d'habiter en ZUS, le type de coordonnées téléphoniques, le mode d'envoi de la lettre-avis et le fait de connaître le nom de la commune de résidence.

Les individus qui habitent en ZUS ainsi que les moins qualifiés ont plus souvent tendance à abandonner en cours d'enquête. Le programme et les résultats de la modélisation logistique sont consultables en annexe.

## **b) Cohérence de la modélisation**

Le modèle permet de prédire le comportement réel des individus dans 59 % des cas. Le pouvoir explicatif du modèle est donc faible.

Sur l'ensemble des 37 675 individus dans le champ ayant commencé un questionnaire, l'estimation de la probabilité moyenne de le terminer est égale à 0,90 (min=0,76 ; max=0,96 ; ET=0,03).

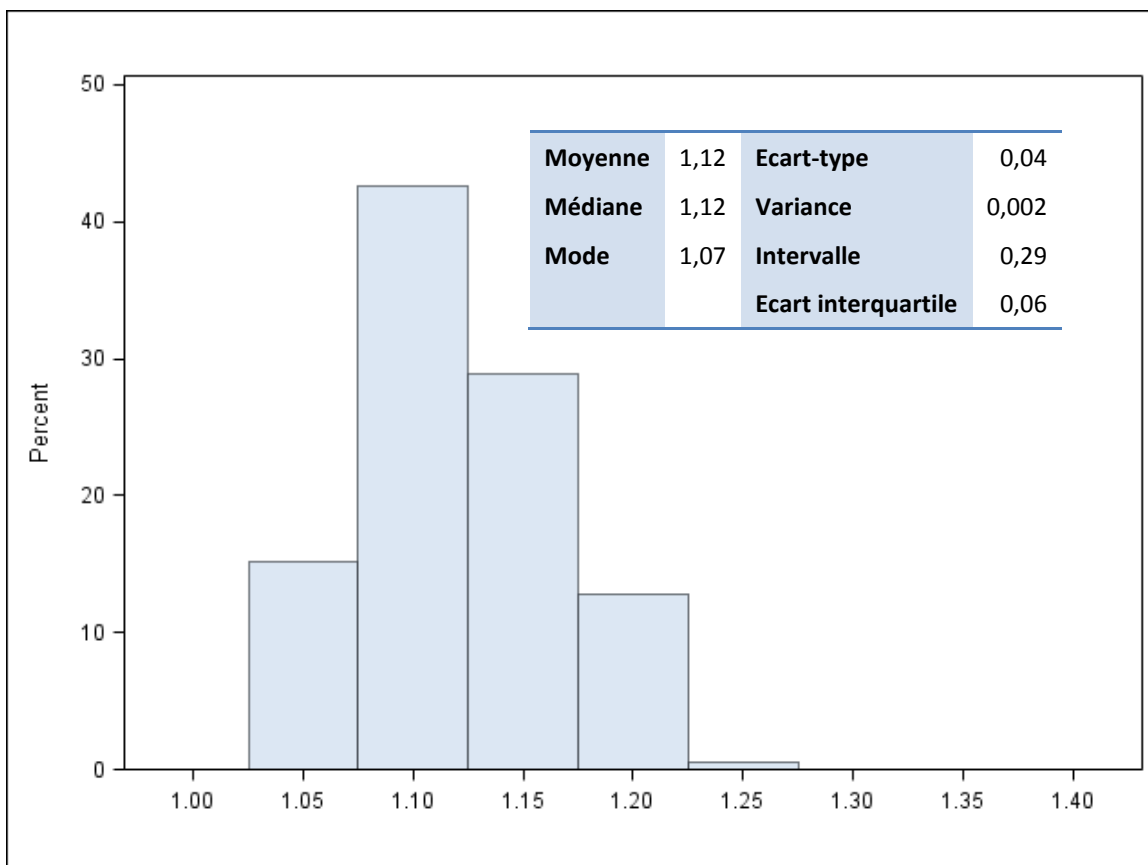
Pour les individus ayant effectivement terminé un questionnaire, le poids relatif est égal à l'inverse de la probabilité de terminer le questionnaire estimée par la procédure logistique.

A l'issue de la phase de constitution de la base de données, un total de 312 questionnaires terminés inexploitable a été supprimé, portant à 33 547 le nombre total de questionnaires exploitables dans le champ du Céreq. Considérant que les questionnaires supprimés se répartissent aléatoirement au sein des questionnaires terminés, un coefficient de correction a été appliqué au poids précédent sur l'ensemble des questionnaires terminés exploitables. La somme des poids relatifs au fait de terminer un questionnaire exploitable pour les individus dans cette situation est égale à 37 680.

### c) Poids relatif au fait de terminer un questionnaire exploitable pour les individus mis à disposition dans la table d'exploitation

Pour les 33 547 individus dans le champ de l'enquête et présents dans la base mise à disposition, il est compris entre 1,05 et 1,33.

**Figure 21 Distribution des poids relatifs au fait de terminer un questionnaire exploitable**



## 7. Le calage sur marges

### a) Principe général

L'enquête Emploi de l'Insee sert de référence dans les publications du ministère de l'Education nationale. Afin d'obtenir des résultats cohérents avec cette source, les données de Génération sont également calées dans ce sens. On utilise la publication « Repères et références statistiques » (RERS 2013) – Fiche 8.23 « Le niveau d'études à la sortie du système éducatif » p. 271 - pour réaliser le calage.

**Tableau 30 - Publication RERS**

### [1] Répartition des sortants de formation initiale en fonction du diplôme le plus élevé

(France métropolitaine, données provisoires)

	Année de sortie de formation initiale					
	2006-2007-2008		2009-2010-2011			
	Ensemble		Ensemble		Hommes	Femmes
	En milliers	En %	En milliers	En %	En %	En %
DEA, DESS, master, doctorat	67	10	85	12	10	15
Écoles supérieures	42	6	37	5	6	4
Licence, maîtrise	73	10	69	10	8	12
Deug, BTS, DUT et équivalent	88	12	83	12	12	11
Paramédical et social	23	3	23	3	1	5
<b>Total diplômés des études supérieures</b>	<b>293</b>	<b>42</b>	<b>297</b>	<b>42</b>	<b>37</b>	<b>47</b>
Baccalauréat général	59	8	61	9	8	10
Baccalauréat technologique, professionnel et assimilé	106	15	131	19	19	19
CAP, BEP ou équivalent	120	17	100	14	16	13
<b>Total diplômés du second cycle du secondaire</b>	<b>285</b>	<b>40</b>	<b>292</b>	<b>42</b>	<b>43</b>	<b>42</b>
Brevet seul	59	8	51	7	8	6
Aucun diplôme	68	10	60	9	12	5
<b>Total brevet et aucun diplôme</b>	<b>127</b>	<b>18</b>	<b>111</b>	<b>16</b>	<b>20</b>	<b>11</b>
<b>Total sortants de formation initiale</b>	<b>705</b>	<b>100</b>	<b>700</b>	<b>100</b>	<b>100</b>	<b>100</b>

Lecture - En moyenne sur 2009, 2010 et 2011, 700 000 jeunes sont sortis de formation initiale. 42 % sont sortis diplômés de l'enseignement supérieur.

On choisit d'utiliser 16 marges de calage, issues du croisement du genre et du niveau du plus haut diplôme obtenu.

Les données du RERS concernent uniquement la France métropolitaine. Le calage est donc réalisé sur les 32 786 individus formés en France métropolitaine. Une correction est effectuée sur les marges de calage car la base de sondage de « Génération 2010 » ne contenait pas les jeunes formés en IUFM (estimés à environ 14 500 sortants dans le champ Céreq pour la France métropolitaine). On estime avant calage à 631 600 le nombre de sortants du système éducatif (hors IUFM) au cours ou à l'issue de l'année scolaire 2009-2010.

Le nombre de sortants avant calage est donc globalement sous-estimé de 7,8 % par Génération par rapport au chiffre RERS.

Au niveau des marges, ce sont les non diplômés et les diplômés de bac général qui sont les plus sous-estimés dans l'enquête du Céreq. Le calage a été effectué à l'aide de la macro-procédure CALMAR<sup>13</sup>, de l'Insee. Le tableau 31 présente les marges utilisées, estimées en compilant les enquêtes trimestrielles de trois années successives, 2009 à 2011.

<sup>13</sup> <https://www.insee.fr/fr/information/2021902>

**Tableau 31 - Les marges de calage pour la France métropolitaine**

	Répartition avant calage	Marges de calage	Poids de calage moyen
Non diplômés Hommes	6,38	10,6	1,67
Non diplômées Femmes	4,17	5,74	1,42
CAP – BEP et équivalent Hommes	9,9	7,94	0,90
CAP – BEP et équivalent Femmes	7,32	6,35	0,94
BAC professionnel et technologique Hommes	10,64	9,78	1,03
BAC professionnel et technologique Femmes	10,02	9,62	1,07
BAC général Hommes	2,66	4,11	1,74
BAC général Femmes	3,62	5,03	1,59
Bac+2 (ensemble), bac+3 de la santé et du social Hommes	6,80	6,91	1,09
Bac+2 (ensemble), bac+3 de la santé et du social Femmes	8,76	8,35	1,03
Bac+3 (hors santé social), bac+4 (ensemble) Hommes	4,79	3,65	0,82
Bac+3 (hors santé social), bac+4 (ensemble) Femmes	5,97	4,70	0,84
Bac+5 et plus Hommes	9,25	7,95	0,92
Bac+5 et plus Femmes	9,72	9,27	1,02
<b>Ensemble</b>	<b>100</b>	<b>100</b>	
<b>Effectifs totaux</b>	<b>631 600</b>	<b>685 500</b>	<b>1,09</b>
<b>Source</b>	Céreq, enquête « Génération 2010 », pondération avant le calage sur marge	Insee Enquêtes emploi 2009, 2010, 2011, (DEPP, L'Etat de l'Ecole, nov. 2013), correction Céreq pour les IUFM	Céreq, enquête « Génération 2010 »

**Tableau 32 - La ventilation selon le plus haut diplôme pour les enquêtes Génération**

<i>(données France métropolitaine)</i>	<b>Génération 2004</b>	<b>Génération 2007</b>	<b>Génération 2010</b>
Non diplômés Hommes	10,53	11,17	10,6
Non diplômées Femmes	6,17	6,83	5,74
CAP – BEP et équivalent Hommes	10,29	10,22	7,94
CAP – BEP et équivalent Femmes	7,00	6,78	6,35
BAC professionnel et technologique Hommes	9,72	9,52	9,78
BAC professionnel et technologique Femmes	8,64	8,00	9,62
BAC général Hommes	2,24	2,20	4,11
BAC général Femmes	3,44	3,29	5,03
Bac+2 (ensemble), bac+3 de la santé et du social Hommes	7,89	7,05	6,91
Bac+2 (ensemble), bac+3 de la santé et du social Femmes	10,69	8,95	8,35
Bac+3 (hors santé social), bac+4 (ensemble) Hommes	4,59	4,50	3,65
Bac+3 (hors santé social), bac+4 (ensemble) Femmes	7,42	7,26	4,70
Bac+5 et plus Hommes	6,18	7,45	7,95
Bac+5 et plus Femmes	5,19	6,79	9,27
<b>Ensemble</b>	<b>100</b>	<b>100</b>	<b>100</b>
<b>Effectifs totaux</b>	<b>737 000</b>	<b>739 000</b>	<b>685 500</b>
<b>Source</b>	Céreq, enquête « Génération 2004 », pondération calée	Céreq, enquête « Génération 2007 », pondération calée	Céreq, enquête « Génération 2010 », pondération calée hors sortants IUFM



## 8. Le récapitulatif

Pour les 33 547 individus dans le champ de l'enquête et répondant au questionnaire, le poids final est compris entre 1,56 et 157,7.

Figure 22- Distribution des poids définitifs

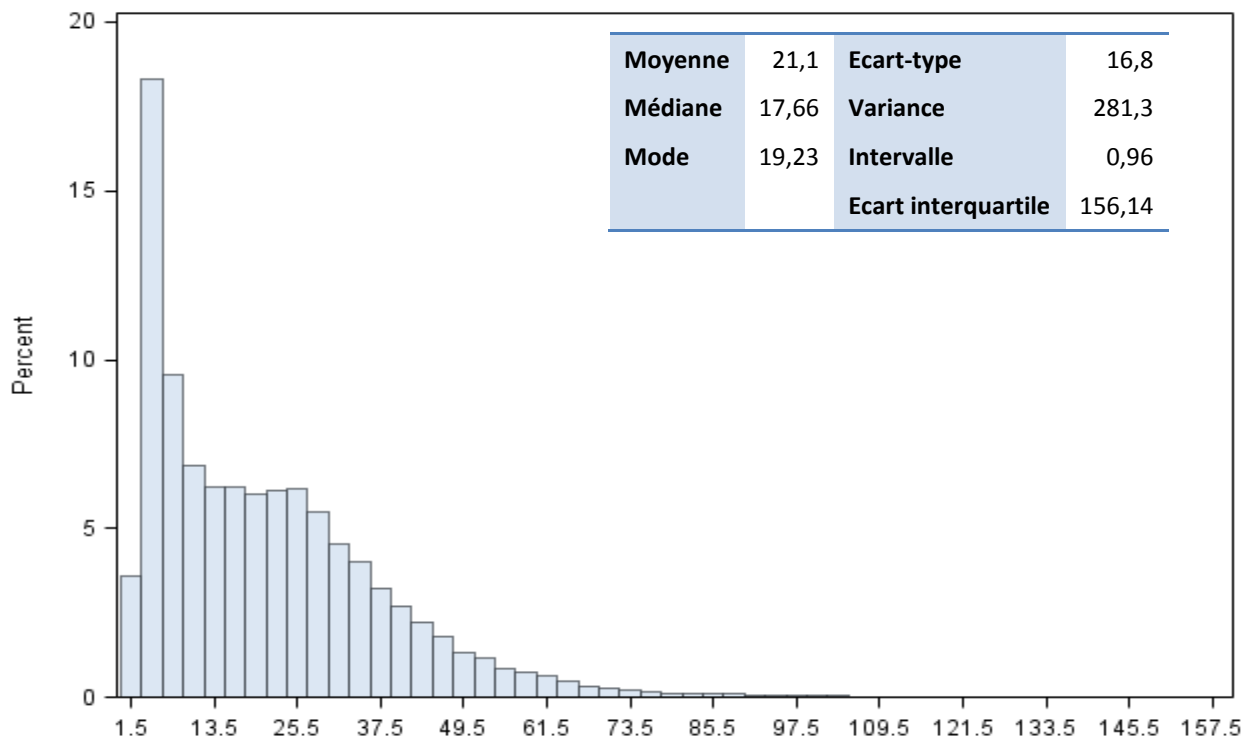


Tableau 33 - Poids définitifs en fonction de la strate de formation

Strate de formation (variable stradef)	Nombre de questionnaires dans le champ	Poids moyen des répondants dans le champ
inconnue	19	31,92
00 – 5 <sup>e</sup> et 4 <sup>e</sup>	54	36,78
01 - autre collège	702	23,36
02 - cap an non term, seconde pro	376	27,14
03 - segpa, prem pro et adap techno	36	18,39
04 - bacpro an non term	1 088	23,26
05 - seconde, 1 <sup>re</sup>	280	31,35
06 - première adaptation bt bac techno	54	14,08
07 - fcil	73	25,44
08 - cap, bep tertiaire an term	2 171	22,97
09 - aide soignant, aux puer, tertiaire post cap, cfa aut niv 5	585	11,49
10 - bep, cap sco indus an term	2 571	26,11
11 - bep cfa indus	321	27,75
12 - bac pro mc bp tertiaire	1 959	27,01

13 - bac pro an term, term pro, bp industriel	802	21,16
14 - bac pro cfa indus	434	20,78
15 - cfa niv 4 bac pro agri indus	485	17,88
16 - bac techno agricole	162	6,61
17 - term bt sti stl	199	26,82
18 - terminale s l es	799	26,65
19 - terminale, post bac tertiaire	1 141	24,70
20 - bac+1 tertiaire lsh droit eco staps	1 485	30,36
21 - bac+1 sciences santé	353	22,55
22 - bma dma dsaa	20	30,13
23 - deug eco aes, decf	35	49,16
24 - deug droit lsh	192	39,74
25 - deug sciences staps	133	17,84
26 - iut 2 secondaire, dnts	81	37,28
27 - dut cfa	26	27,57
28 - drjs mon educ	697	6,80
29 - bts dut tertiaire	1 463	29,90
30 - bts indus	1 030	26,72
31 - ass serv soc educ	871	4,65
32 - conseiller eco so fami	167	8,00
33 - bac+2 santé	2 123	11,25
34 - licence pro univ	888	20,20
35 - iufm	16	61,61
36 - licence lsh	411	24,06
37 - licence santé, niv 3 cfa	102	23,36
38 - licence droit eco	337	20,20
39 - licence sciences staps	362	10,09
40 - m1 universitaire	1 246	14,64
41 - école art	286	30,37
42 - autres écoles	113	27,92
43 - niv 2 univ priv cfa	243	22,75
44 - niv 1 univ priv cfa	97	23,23
45 - master sante	43	21,90
46 - master lsh droit eco	1 340	28,97
47 - master sciences staps santé vétérinaire	775	16,84
48 - écoles commerce	560	27,14
49 - écoles ingénieur	949	23,21
50 - docteur médecine pharmacie	283	34,99
51 - efp	262	7,25
52 - thèses	2 247	4,25
<b>total</b>	<b>33 547</b>	<b>21,10</b>

## 9. La pondération des extensions

Les traitements pour obtenir les pondérations des extensions de champ sont présentés en annexe 10. Ces traitements concernent les trois extensions de champ du sport, de la santé et de la fonction publique.

## VIII. Une estimation de la variance par calcul direct

Nous avons procédé à une estimation de la variance par calcul direct, en composant les formules associées respectivement aux étapes de tirage d'échantillon, à la correction de la non-réponse et à l'étape de calage. Cette partie a été réalisée avec l'appui méthodologique de la division sondage de l'Insee. Comme l'enquête Génération mobilise essentiellement des données qualitatives, ce seront les variances des proportions qui seront calculées à partir des variables linéarisées.

Le choix de l'estimation de la variance par calcul direct a été fait pour des raisons de coût et des raisons méthodologiques. La base de sondage comporte en effet plus d'un million de lignes et le temps de calcul de la macro cube croit rapidement avec le nombre d'observations. De ce fait, Les tirages d'échantillons équilibrés par strates avaient été effectués en mobilisant plusieurs PC au sein du département. Il aurait été très lourd de procéder à un calcul de précision par répliquions sur une population fictive de plus d'un million d'observations, étant donné que le minimum de répliquions pour obtenir un écart type est de l'ordre de 30. Nous aurions pu calculer les variances par bootstrap non pas sur une population fictive de la même taille que celle de la base de sondage mais sur une population pondérée. Cependant, nous n'avons pas de garantie sur la validité de cette méthode.

### 1. Estimation de la variance due à l'échantillonnage équilibré

L'intérêt du sondage équilibré est de diminuer la variance des variables d'intérêts, la variance étant la variance des résidus sur les variables d'équilibrage. De plus, lorsque l'équilibrage est exact, la structure de l'échantillon correspond exactement à la structure du système des variables auxiliaires spécifiées. La variance du sondage équilibré est estimée par l'approximation de variance de Deville Tillé. Etant donné que le sondage a été stratifié par région et selon des catégories de formation, la variance totale sera la somme sur chaque strate des variances. De plus, comme la population a été restreinte à l'ensemble  $S_{restreint}$ , ce seront uniquement des variances calculées sur ces strates restreintes du supérieur qui seront sommées en s'intéressant en particulier au domaine  $S_d$ .

Cette formule de variance considère que le tirage équilibré est un tirage poissonnien réalisé sous les contraintes d'équilibrage. Sa variance est celle d'un sondage poissonnien avec des expressions particulières des paramètres  $C_{k,h}$ . L'application de l'approximation de variance de Deville Tillé pour estimer la variance de l'estimateur d'Horvitz-Thomson du total de la variable d'intérêt Y sur la strate h, notée  $\hat{V}_{DT}(\hat{Y}_\pi^h)$ , est donnée par :

$$\hat{V}_{DT}(\hat{Y}_\pi^h) = \sum_{k \in S_h} c_{k,h} (y_k - \tilde{y}_{k,h})^2$$

où  $\tilde{y}_{k,h} = x'_k \left( \sum_{l \in S_h} c_l x_l x_l' \right)^{-1} \sum_{l \in S_h} c_l x_l y_l$  avec  $x'_k = (x_{k1}, \dots, x_{kq})$  et  $c_{k,h} = \frac{(1 - \pi_k) n_h}{\pi_k^2 n_h - q}$ .

La quantité  $(y_k - \tilde{y}_{k,h})$  correspond en fait au résidu de la régression dans  $S_h$  de la variable d'intérêt Y sur les variables de calage X, pondérée par  $c_{k,h}$  : elle se calcule donc très facilement à l'aide d'une proc reg, et la quantité  $\hat{V}_{DT}(\hat{Y}_\pi^h)$  facilement. On trouvera en annexe un programme SAS générique réalisant ce calcul.

La variance de l'estimateur d'Horvitz-Thompson du total d'une variable d'intérêt Y liée au seul aléa d'échantillonnage est donc estimée par :

$$\hat{V}_{DT}(\hat{Y}_\pi) = \sum_h \hat{V}_{DT}(\hat{Y}_\pi^h) = \sum_h \sum_{k \in S_h} c_{k,h} (y_k - \tilde{y}_{k,h})^2$$

## 2. Estimation de la variance en prenant en compte le processus de non-réponse

L'approximation de la variance avec non réponse, se fait en se basant sur la variance obtenue dans un sondage à deux degrés, où la phase de non réponse est considérée comme le deuxième degré de la phase de sondage. En partant de la formule de décomposition de la variance on a :

$$V(\hat{T}) = E_1(V_{2/1}\hat{T}) + V_1(E_{2/1}\hat{T})$$

Où les indices 1 et 2 désignent les premier et second degrés de tirage.

$T_i$  est le total de la variable Y dans les unités secondaires. La formule de Rao permet d'exprimer la variance du total comme la somme d'un terme quadratique et de la somme des variances des  $T_i$

$$\hat{V}(\hat{T}) = Q(\hat{T}_1, \dots, \hat{T}_m) + \sum_{i \in S_1} \left( \frac{1}{\pi_i^2} - q_i \right) \hat{V}(\hat{T}_i)$$

La variance du premier terme est la variance due à la partie équilibrage, le deuxième terme est la correction à apporter à l'estimation de la variance en prenant en compte l'ajout de variance dû à la non-réponse.

Dans notre cas, la deuxième phase correspond au processus de non-réponse, une unité secondaire correspond donc à un individu. La variable d'intérêt est la variable  $T_i = y_i$ , le total sur l'unité secondaire i. elle

est estimée par  $\hat{T}_i = \frac{y_i \prod_{i \in R}}{p_i}$ , avec  $p_i$  la probabilité de répondre de l'individu i et  $I_{i \in R}$  l'indicatrice

d'appartenance à la population des répondants.

L'estimation de la variance prenant en compte l'aléa de sondage et la non-réponse est donnée par :

$$\hat{V}(\hat{Y}_{\pi, \text{CNR}} = \sum_{k \in S} \frac{y_k}{\pi_k p_k}) = \sum_h \sum_{k \in S_h} c_{k,h} (z_k - \tilde{z}_{k,h})^2 + \sum_{k \in S} \left( \frac{1}{\pi_k^2} - q_k \right) (1 - p_k) z_k^2$$

avec  $z_i = \frac{y_i \prod_{i \in R}}{p_i}$ .

Le premier terme correspond à la variance d'équilibrage et le deuxième au terme additionnel dû à la prise en compte de la non-réponse.

## 3. Estimation de la variance de l'estimateur à l'issue de la phase de calage

Lorsque la taille de l'échantillon augmente, l'estimateur calé converge vers l'estimateur par la régression. De ce fait, il est de biais négligeable, et sa variance est celle des résidus obtenus par la régression des variables d'intérêts sur les variables de calage. De manière générale, si on a un estimateur calé  $T_W$ , la variance de ce dernier est donné par :

$$V(\hat{T}_W) \approx V\left(\sum_{k \in S} d_k \varepsilon_k\right)$$

Où  $\varepsilon_k = y_k - \sum_{\alpha=1}^K \hat{b}_\alpha X_k^\alpha$ , est le résidu de  $y$  sur les variables explicatives  $X^\alpha$  ( $\alpha = 1, 2, \dots, K$ ), et les  $d_k$  sont les poids des observations.

On va calculer la variance à partir des résidus de la variable d'intérêt sur les variables de calage avec comme pondération les poids avant calage prenant en compte le poids de tirage et le poids lié à la correction de la non-réponse. L'estimation de variance en prenant en compte les trois étapes de tirage d'échantillon, de correction de la non-réponse et de calage est donné par :

$$\hat{V}(\hat{Y}_{\text{calé}}) = \sum_h \sum_{k \in S_h} c_{k,h} (\zeta_k - \tilde{\zeta}_{k,h})^2 + \sum_{k \in S} \left( \frac{1}{\pi_k^2} - q_k \right) (1 - p_k) \zeta_k^2$$

$$\text{Où } \zeta_i = \frac{\varepsilon_i \mathbb{I}_{i \in R}}{p_i}$$

#### 4. Calcul de la variable linéarisée

Les variables d'intérêts de l'enquête Génération sont essentiellement des variables qualitatives. Les statistiques sont donc des ratios. Cette statistique n'est pas linéaire et l'approximation de la variance a été faite en calculant la linéarisée de la proportion  $P$ . Soit la modalité  $m_v$  d'une variable  $v$  donnée. On va calculer la linéarisée de la proportion des sortants satisfaisant la modalité  $m_v$  sur le nombre de sortants concernés par la question  $v$ .

$$P = \frac{\sum_{k \in U} \mathbb{I}_{i \in m_v}}{\sum_{k \in U} \mathbb{I}_{i \in v}} = \frac{M_v}{N_v}, \text{ proportion calculée sur la population totale } U.$$

On considère une fonction de totaux dérivable  $f(t_1, \dots, t_q)$ , avec  $t_1$  à  $t_q$  les totaux des variables  $y_1$  à  $y_q$

La variable linéarisée  $u_k$  est définie par :

$$u_k = \sum_{i=1}^q \frac{\partial f}{\partial t_i}(t_1, \dots, t_q) y_{ki}$$

On considère une fonction de totaux correspondant au ratio  $\theta = \frac{t_1}{t_2}$

La variable linéarisée est alors :

$$u_k = \frac{\partial f}{\partial t_1} y_1 + \frac{\partial f}{\partial t_2} y_2$$

$$\text{Soit } u_k = \frac{1}{t_2} y_1 - \frac{t_1}{t_2^2} y_2$$

Avec  $y_1$  = indicatrice  $\mathbb{I}_{i \in m_v}$  égale 1 si l'observation satisfait la modalité  $m_v$

$y_2$  une indicatrice  $\mathbb{I}_{i \in v}$  d'appartenance au champ de la question

$$\text{Donc } u_k = \frac{1}{N} (\mathbb{I}_{i \in m_v} - P \mathbb{I}_{i \in v})$$

Qui est estimé par l'estimateur par substitution

$$\hat{u}_k = \frac{1}{\hat{N}} (\mathbb{I}_{i \in m_v} - \hat{P} \mathbb{I}_{i \in v})$$

Pour chaque estimation, on fait une estimation sur domaine. Le domaine n'est pas uniquement défini par rapport à  $S_d$  il faut aussi tenir compte du champ de la question. Si l'observation appartient à  $S_d$  et

appartient au champ de la question alors  $\hat{u}_k = \frac{1}{\hat{N}} (\mathbb{I}_{i \in m_v} - \hat{P} \mathbb{I}_{i \in v})$ , sinon  $\hat{u}_k = 0$

Ces observations qui n'appartiennent pas au domaine participent tout de même au calcul de la variance. En effet, dans les étapes de régression sur les variables de calage, on peut avoir des valeurs qui ne soient plus égales à zéro, et dans ce cas participent à l'augmentation de la variance. Ces termes peuvent être très gênants sur les domaines extrêmement petits, où la variance peut être rédhibitoire.

## 5. Exemple de variance obtenue

Le tableau ci-dessous présente les intervalles de confiance des estimateurs de proportion de la variable dichotomisé SITDE (situation à la date d'enquête) sur la France entière.

**Tableau 34 - Variance des proportions pour la variable « Situation à la date d'enquête »**

Situation à la date d'enquête	CATI
emploi	67,52 ( $\pm 0,45$ )
chômage	19,87 ( $\pm 0,39$ )
inactivité	4,04 ( $\pm 0,20$ )
formation	2,47 ( $\pm 0,16$ )
reprise d'étude	6,09 ( $\pm 0,23$ )

## IX. Publications récentes

Les premières valorisations de l'enquête 2013 auprès de la « Génération 2010 » sont listées ci-dessous :

- *Céreq Bref* n°319 de premiers résultats : « Face à la crise, le fossé entre niveaux de diplôme se creuse » (Barret, Ryk, Volle ; avril 2014) (8 pages)
- 2 *Céreq Brefs* thématiques :
  - o Bref n°322 : « Sortants du supérieur : la hausse du niveau de formation n'empêche pas celle du chômage » (Ménard ; septembre 2014)
  - o Bref n°335 : « CAP-BEP : des difficultés d'insertion encore aggravées par la crise" (Ilardi, Sulzer ; mai 2015).
- Ouvrage collectif du Deeva « Quand l'école est finie. Premiers pas dans la vie active » (23 octobre 2014).
- Hors-série pratique de la revue Alternatives économiques (n°71, « Entrer dans la vie active », janvier 2015).

Les retombées médiatiques et institutionnelles de la conférence de presse du 8 avril 2014 sur le Bref de premiers résultats ont été nombreuses ... Plusieurs reprises dans la presse nationale et régionale avec notamment la une du Monde mais aussi des reprises radio et télé. Le ministre de l'Education, Benoît Hamon, tout juste nommé, a cité l'étude dans son discours d'introduction. La conférence de presse sur la sortie de l'ouvrage « Quand l'école est finie » s'est faite en présence de la ministre de l'Education, Najat Vallaud-Belkacem. On peut noter également plusieurs auditions institutionnelles (Igas, auditions parlementaires, ...).

Cela dénote un intérêt majeur du thème de l'insertion professionnelle des jeunes dans le débat public et institutionnel. Le dispositif des enquêtes Génération est ainsi bien ancré dans l'offre de statistique publique, et forme un des piliers de l'offre de production statistique du Céreq.



## X. Perspectives pour les enquêtes futures

Le financement des enquêtes Génération est assuré jusqu'en 2017 (3<sup>e</sup> interrogation de la « Génération 2010 »). Au-delà, compte tenu des contraintes budgétaires, il est nécessaire de rénover le dispositif. Un groupe de travail est constitué, il a pour objectif de dresser la liste des pistes techniquement, scientifiquement et budgétairement valides, en présentant systématiquement les avantages et inconvénients. Si ce groupe a vocation à réfléchir globalement à l'introduction d'un nouveau mode de collecte en s'appuyant sur l'expertise méthodologique de « Génération 2010 », il a également pour objectif de proposer une nouvelle expérimentation, multimode cette fois, pour la ré-interrogation de 2015.

La piste des enquêtes multimodes n'est pas la seule manière de réduire les coûts. Toutes les autres pistes seront explorées : le rythme des interrogations (alternance petite et grande génération, diminution/suppression des réinterrogations, espacement entre les différentes enquêtes, ...), la réduction de la taille de l'échantillon, de la longueur du questionnaire.

Une réflexion sur le questionnaire et sur le champ de l'enquête sera menée en parallèle, à partir d'une recension des expériences nationales et internationales. En effet, une rénovation du dispositif est une opportunité pour réinterroger la notion même d'insertion comme d'autres catégories de compréhension plus ou moins essentialisées (comme celles de « sortant » ou de « débutant »). D'une part, les frontières entre la formation et l'activité professionnelle tendent à s'effacer (alternance, stages, travail en cours d'études...) et les parcours de sortie de l'appareil scolaire ou de l'enseignement supérieur pour entrer dans la vie active sont de moins en moins linéaires, de plus en plus longs, complexes et instables. En outre, la nature des indicateurs d'insertion s'avère de plus en plus problématique : ceux-ci doivent-ils prendre en compte le dernier ou le diplôme le plus élevé ou bien mesurer l'ensemble du parcours scolaire ? De même, le taux d'emploi ne présente-t-il pas des insuffisances au regard de la qualité de l'emploi ou des emplois successifs occupés ? Par ailleurs, les travaux réalisés sur les inégalités (nationale, sociologique, de genre) ne montrent-ils pas que derrière les caractéristiques individuelles et les parcours d'études, les effets de l'origine sociologique sur l'insertion demeurent ?

Parmi les pistes de financement, un développement des partenariats avec les régions (en particulier lorsqu'elles auront fusionnées) devra être examiné. L'ARF ou le réseau des Carif-Oref seront contactés en ce sens. En effet, avec la régionalisation de la politique publique de formation professionnelle, d'orientation et d'apprentissage, le rôle des régions se trouve renforcé. Il est ainsi légitime de penser que les besoins d'analyse régionale sur l'insertion des jeunes seront grandissants.

Parmi les autres pistes de financement, il sera nécessaire d'examiner les possibilités de financement par les partenaires sociaux, par les fonds pour la recherche, par les fonds européens et internationaux. Des contacts seront également pris avec les autres producteurs d'enquêtes d'insertion (universités, conférence des grandes écoles, ...) Quelles articulations dès lors concevoir entre celles-ci et les enquêtes Génération et leurs extensions ?

La réflexion sur les financements est étroitement liée à celle sur l'évolution du dispositif. Chaque financeur ayant ses propres problématiques et son propre champ d'analyse.

## XI. Annexes

### Annexe 1 – Table des illustrations

#### Les tableaux

Tableau 1 - Détail des enquêtes Génération et effectifs des répondants .....	7
Tableau 2- Extension de l'enquête Génération.....	10
Tableau 3 - Détail de la réception des fichiers par type d'établissement (Hors bases collectées via les rectorats) .....	21
Tableau 4 - Nature des fichiers collectés .....	22
Tableau 5 - Effectifs dans la base de sondage par type d'établissement .....	24
Tableau 6 - Estimation de la couverture des universités par niveau (hors IUFM) .....	26
Tableau 7 - Estimation de la couverture par discipline.....	26
Tableau 8 - Estimation de la couverture du champ CFA .....	27
Tableau 9 - Estimation de la couverture du champ sport .....	28
Tableau 10 - Estimation de la couverture du champ Santé social .....	29
Tableau 11 - Estimation de la couverture des écoles d'ingénieurs et des écoles de commerce.....	29
Tableau 12 - Taux de couverture établissement des autres établissements.....	30
Tableau 13 - Tableau récapitulatif des taux de couverture .....	31
Tableau 14 - Evolution du nombre de numéros de téléphone selon l'enquête .....	32
Tableau 15 - Fréquence de numéros de téléphone selon le type.....	32
Tableau 16 - Résultats RNVP .....	46
Tableau 17 - Résultats des recherches des déménagés.....	47
Tableau 18 : Résultats des recherches de mails.....	48
Tableau 19 - Bilan de la phase A (prestataire 1) .....	50
Tableau 20 - Bilan de la phase D prestataire 1.....	50
Tableau 21 - Bilan de la phase B et C prestataire 1.....	51
Tableau 22 - Bilan chiffré de l'enrichissement .....	52
Tableau 23 - Répartition du nombre de lettres avis envoyées selon le mode et la source .....	53
Tableau 24 - Les règles d'appel détaillées.....	61
Tableau 25 – La codification du contrat de travail.....	73
Tableau 26 - Définition du modèle type sur les salaires (avec évolution inter-enquête).....	77
Tableau 27 – Revalorisation du SMIC entre juin 2010 et juillet 2013 .....	78
Tableau 28 – Normes de rémunérations des contrats de professionnalisation .....	79
Tableau 29 - Normes de rémunérations des contrats d'apprentissage.....	79
Tableau 30 - Publication RERS.....	92
Tableau 31 - Les marges de calage pour la France métropolitaine.....	93
Tableau 32 - La ventilation selon le plus haut diplôme pour les enquêtes Génération .....	94
Tableau 33 - Poids définitifs en fonction de la strate de formation .....	95
Tableau 34 - Variance des proportions pour la variable « Situation à la date d'enquête » .....	101

## Les figures (graphiques, schémas)

Figure 1 - Calendrier des enquêtes Génération .....	7
Figure 2 – Le calendrier professionnel .....	12
Figure 3 - Schéma organisationnel de la collecte.....	19
Figure 4 Les appariements entre bases de données / Estimation du taux de couverture des individus de la base de sondage.....	25
Figure 5 - Distribution de la probabilité estimée de réponse au questionnaire filtre au sein de l'échantillon de « Génération 2004 ».....	37
Figure 6 - Distribution de la probabilité estimée de réponse au questionnaire filtre au sein de l'échantillon de « Génération 2007 ».....	38
Figure 7 - Distribution de la probabilité estimée au sein de la base de sondage de « Génération 2010 », à partir des résultats de « Génération 2004 » et « Génération 2007 ».....	38
Figure 8 - Distribution de la probabilité estimée retenue au sein de la base de sondage de « Génération 2010 » .....	39
Figure 9 - Répartition des déménagés .....	47
Figure 10 - Enrichissements des coordonnées via le numéro vert .....	54
Figure 11 - Enrichissement des coordonnées via le Cawi .....	54
Figure 12 - Calendrier et organisation générale avec le prestataire .....	59
Figure 13 : Nombre d'enquêtes réalisées selon la période et le nombre de télé-enquêteurs .....	60
Figure 14 : durée moyenne du questionnaire selon la semaine d'enquête .....	60
Figure 15 - Distribution de la durée de passation du questionnaire.....	61
Figure 16 - Gestion des salaires.....	75
Figure 17 Distribution des poids de correction liés au défaut de couverture de la base de sondage .....	84
Figure 18 Distribution des poids d'échantillonnage.....	85
Figure 19 - Distribution des poids de contact .....	87
Figure 20 - Distribution des poids relatifs au fait d'accepter de répondre .....	89
Figure 21 Distribution des poids relatifs au fait de terminer un questionnaire exploitable.....	91
Figure 22- Distribution des poids définitifs .....	95

## Les encadrés

Encadré 1 - Le Centre d'études et de recherches sur les qualifications .....	6
Encadré 2 - Une enquête de la statistique publique.....	9
Encadré 3 - Les informations demandées sur les séquences du calendrier mensuel d'activité.....	12
Encadré 4 - La base centrale des établissements (BCE) .....	18
Encadré 5 - Message laissé sur les répondeurs des interlocuteurs : .....	62
Encadré 6 - SICORE Environnement diplôme et spécialité (millésime 2013).....	68
Encadré 7 - SICORE Environnement Activité (millésime 2012) .....	71
Encadré 8 - SICORE Environnement PCS 2013 .....	72

## Annexe 2 – Lettres avis de contact avec les jeunes

### Lettre envoyée par courrier POSTAL



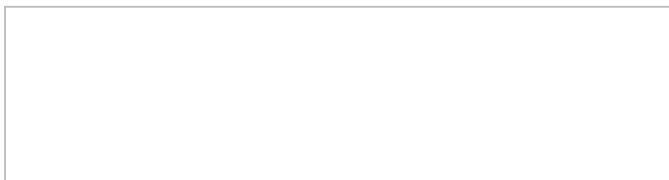
REPUBLIQUE FRANÇAISE

Centre d'études  
et de recherches  
sur les qualifications

Céreq

MINISTÈRE DE L'ÉDUCATION  
NATIONALE

MINISTÈRE DU TRAVAIL  
DE L'EMPLOI, DE LA FORMATION  
PROFESSIONNELLE ET DU DIALOGUE SOCIAL



Madame, Mademoiselle, Monsieur,

Le Centre d'études et de recherches sur les qualifications (Céreq) mène une grande enquête nationale pour connaître le devenir professionnel des anciens élèves ou étudiants issus des différentes filières de formation.

Cette enquête est utile aux acteurs publics de l'orientation et de l'insertion mais aussi pour les jeunes et leurs familles, par exemple lorsqu'ils doivent réfléchir à leur orientation scolaire.

Pour cela, nous interrogeons, à partir de ce mois-ci, un échantillon parmi les jeunes sortis du système éducatif en 2009-2010. Vous avez été choisi pour participer à cette enquête et votre collaboration est essentielle. Aussi, nous vous contacterons par téléphone dans les prochaines semaines.

L'objectif de cet entretien sera de décrire votre parcours professionnel depuis 2010. Il durera vingt minutes environ et pourra varier selon la diversité de votre parcours professionnel.

Vous trouverez ci-joint un calendrier qui vous aidera à reconstituer votre parcours professionnel depuis 2010. Nous vous invitons à le remplir et à le conserver afin de l'utiliser lors de notre entretien téléphonique.

Vos réponses serviront uniquement à produire des statistiques : elles resteront strictement anonymes et confidentielles.

Je vous remercie très vivement de votre participation à cette enquête.

Le directeur du Céreq  
Alberto Lopez

Merci de nous communiquer par internet votre numéro de téléphone en vous connectant à l'adresse suivante <http://www.cereq.fr/g10>, en utilisant **IDENT** et **MDP**.

Vous pouvez également nous le communiquer au numéro vert suivant (appel gratuit depuis un téléphone fixe, en semaine entre 9h et 20h30, le samedi entre 10h et 18h) : **0 800 97 00 55** en mentionnant la référence.

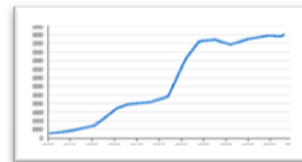
Vous pouvez aussi utiliser ce numéro pour nous indiquer les moments les plus favorables pour vous joindre.

Vu l'avis favorable du Conseil National de l'Information Statistique, cette enquête est reconnue d'intérêt général et de qualité statistique sans avoir de caractère obligatoire. Visa n°2013X705AU du ministère de l'Éducation nationale et du ministère du Travail, de l'Emploi, de la Formation Professionnelle et du Dialogue Social et du Ministère de l'économie et des finances, valable pour l'année 2013. En application de la loi n°51-711 du 7 juin 1951, les réponses à ce questionnaire sont protégées par le secret statistique et destinées à la production de statistiques par le Céreq. La loi n° 78-17 du 6 janvier 1978 modifiée relative à l'informatique, aux fichiers et aux libertés, s'applique aux réponses faites à la présente enquête. Elle garantit aux personnes concernées un droit d'accès et de rectification pour les données les concernant. Vous pouvez également, pour des motifs légitimes, vous opposer au traitement des données vous concernant. Ce droit peut être exercé auprès du Céreq, 10 place de la Joliette, 13567 Marseille Cedex 02, ou bien par un mail adressé au correspondant informatique et libertés du Céreq à l'adresse suivante :

[cil-generation@cereq.fr](mailto:cil-generation@cereq.fr) en mentionnant dans l'objet « Génération 2010 » suivi de votre identifiant apparaissant sur la première page de ce courrier. De plus, si vous êtes mineur, le titulaire de l'autorité parentale devra donner son accord pour que vous puissiez participer à cette étude.

### POURQUOI VOUS INTERROGE-T-ON ?

Cette enquête cherche à décrire les parcours et les situations professionnelles des jeunes quels que soient le diplôme préparé ou la formation suivie en 2009-2010. Les résultats permettront d'améliorer l'information des jeunes et de leur famille au moment de l'orientation scolaire.



### ...PAR TELEPHONE

Afin que l'entretien se déroule de manière fluide et confortable, et pour assurer une bonne compréhension de vos réponses, l'utilisation d'un téléphone fixe est préférable.



### LA CONFIDENTIALITE DES DONNEES



Vos réponses au questionnaire sont couvertes par le secret statistique et ne donneront lieu qu'à des exploitations anonymes. Vos coordonnées seront systématiquement supprimées au moment de l'exploitation de l'enquête.

### QU'EST-CE QUE LE CEREQ ? QUEL EST SON ROLE ?



Le Centre d'études et de recherches sur les qualifications est un établissement public sous tutelle du ministère de l'Education nationale et du ministère du Travail.

Pôle public d'expertise au service des acteurs de la formation et de l'emploi, il allie production de statistiques, recherches, études et accompagnement d'actions.

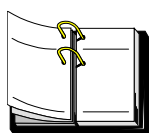


## POUR PREPARER NOTRE ENTRETIEN TELEPHONIQUE

### ☞ Nous vous suggérons afin de faciliter l'entretien, de remplir le calendrier professionnel de 2010 à 2013 (voir au dos) :

- en vous aidant de vos CV, contrats de travail et bulletins de salaire
- en faisant appel aux événements qui ont marqué votre vie afin de vous repérer dans le temps

### ➤ Voici quelques repères chronologiques pour la période 2010 – 2013 :



- Juin 2010... Coupe du monde de football en Afrique du Sud : Grève des joueurs de l'équipe de France à Knysna
- Juillet 2011... Décès de la chanteuse Amy Winehouse
- Sept 2011... Coupe du monde de Rugby : La France est battue en finale par la Nouvelle Zélande
- Mai 2012... Election présidentielle en France : F. Hollande nouveau Président de la République
- Aout 2012... Jeux Olympiques à Londres

### ☞ Quelques règles :

- Les diverses situations sont exclusives, c'est-à-dire que pour une période donnée, vous ne pouvez décrire qu'une seule situation.
- Si vous avez vécu plusieurs situations différentes pendant une même période, privilégiez toujours **le travail** sur les autres situations.
- Si, pendant un mois donné, vous avez travaillé trois semaines et le reste du mois vous étiez en recherche d'emploi, alors considérez le mois entier comme une période de travail.

### ☞ Quelques précisions sur les situations :

- « **En Intérim** » : vous avez effectué plusieurs missions d'intérim dans des entreprises différentes durant une période donnée. Si vous étiez intérimaire dans une seule et même entreprise, classez-vous dans la situation « **Dans une même Entreprise** »
- « **En reprise d'études** » : utiliser cette ligne uniquement si vous avez repris vos études à **temps plein dans un établissement scolaire ou universitaire**.
- « **En formation** » : Cette rubrique ne concerne que les périodes de formation **hors contrat de travail**.

ix. Pour plus d'informations et pour retrouver des résultats des études précédentes,  
x. Vous pouvez consulter le site du Céreq : [www.cereq.fr](http://www.cereq.fr)

## Lettre envoyée par courrier ELECTRONIQUE



Centre d'études  
et de recherches  
sur les qualifications

Céreq

MINISTÈRE DE L'ÉDUCATION  
NATIONALE

MINISTÈRE DU TRAVAIL,  
DE L'EMPLOI, DE LA FORMATION  
PROFESSIONNELLE ET DU DIALOGUE SOCIAL

Madame, Monsieur,

Le Centre d'études et de recherches sur les qualifications (Céreq) mène **une grande enquête nationale** pour connaître le devenir professionnel des anciens élèves ou étudiants issus des différentes filières de formation.

Cette enquête est utile aux acteurs publics de l'orientation et de l'insertion mais aussi pour les jeunes et leurs familles, par exemple lorsqu'ils doivent réfléchir à leur orientation scolaire.

Pour cela, nous interrogeons, à partir du mois d'avril prochain, un échantillon parmi les jeunes sortis du système éducatif en 2009-2010. **Vous avez été choisi pour participer à cette enquête et votre collaboration est essentielle. Aussi, nous vous contacterons par téléphone dans les prochaines semaines.**

**<Si votre numéro de téléphone n'est plus le : 06xxxxxxxx >  
Ou <Nous ne disposons pas de vos coordonnées téléphoniques.>**

*Vous pouvez nous communiquer par internet votre numéro de téléphone en vous connectant à l'adresse suivante :*

[http://na2.voxco.com/Media/1004/CEREQ\\_actualisation.htm](http://na2.voxco.com/Media/1004/CEREQ_actualisation.htm)

*Vous pouvez aussi nous contacter au numéro vert suivant (appel gratuit depuis un téléphone fixe, en semaine entre 9h et 20h30, le samedi entre 10h et 18h) : 0 800 97 00 55 en mentionnant la référence : IDENT*

*Vous pouvez utiliser ce numéro vert pour nous indiquer les moments les plus favorables pour vous joindre.*

L'objectif de cet entretien sera de décrire votre parcours professionnel depuis 2010. Il durera vingt minutes environ et pourra varier selon la diversité de votre parcours professionnel.

Vous trouverez en cliquant [ici](#) un calendrier qui vous aidera à reconstituer votre parcours professionnel depuis 2010. **Nous vous invitons à le remplir puis à l'imprimer, afin de l'utiliser lors de l'entretien téléphonique.**

Des informations complémentaires sur notre étude sont disponibles [ici](#).

Vos réponses serviront uniquement à produire des statistiques : elles resteront strictement anonymes et confidentielles.

Je vous remercie très vivement de votre participation à cette enquête.

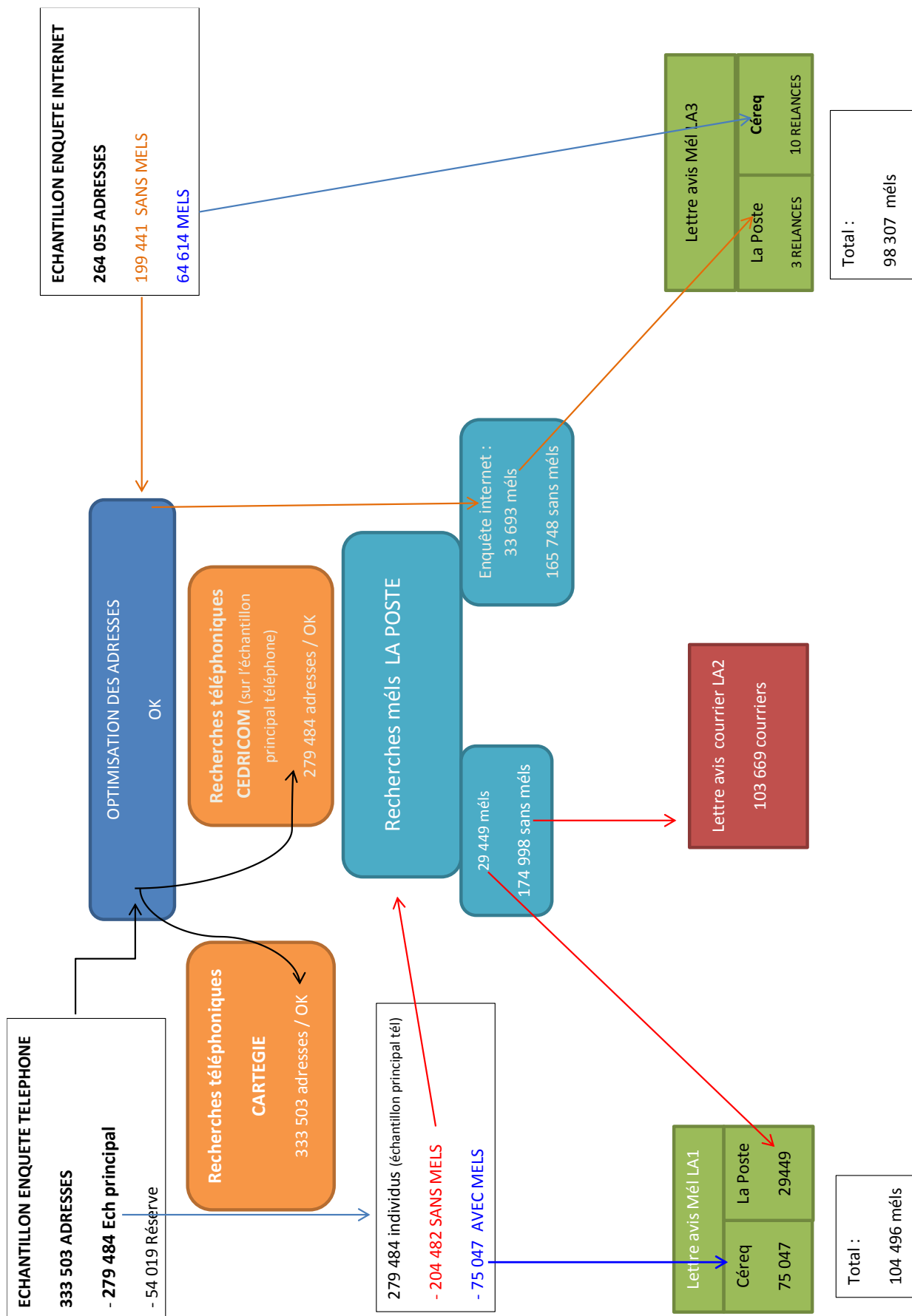
Le directeur du Céreq  
Alberto Lopez



Vu l'avis favorable du Conseil National de l'Information Statistique, cette enquête est reconnue d'intérêt général et de qualité statistique sans avoir de caractère obligatoire. Visa n° 2013X705AU du ministère de l'Education nationale et du ministère du Travail, de l'Emploi, de la Formation Professionnelle et du Dialogue Social et du Ministre de l'économie et des finances, valable pour l'année 2013. En application de la loi n° 51-711 du 7 juin 1951, les réponses à ce questionnaire sont protégées par le secret statistique et destinées à la production de statistiques par le Céreq. La loi n° 78-17 du 6 janvier 1978 modifiée relative à l'informatique, aux fichiers et aux libertés, s'applique aux réponses faites à la présente enquête. Elle garantit aux personnes concernées un droit d'accès et de rectification pour les données les concernant. Vous pouvez également, pour des motifs légitimes, vous opposer au traitement des données vous concernant. Ce droit peut être exercé auprès du Céreq, 10 place de la Joliette, 13567 Marseille Cedex 02, ou bien par un mail adressé au correspondant informatique et libertés du Céreq à l'adresse suivante : [cil-generation@cereq.fr](mailto:cil-generation@cereq.fr) en mentionnant dans l'objet « Génération 2010 » suivi de votre identifiant apparaissant sur la première page de ce courrier. De plus, si vous êtes mineur, le titulaire de l'autorité parentale devra donner son accord pour que vous puissiez participer à cette étude.

Pour ne plus recevoir de courrier électronique de la part du Céreq cliquer [ici](#).

## Organisation de l'envoi des lettres avis



## Annexe 3 – Nomenclature des diplômes

### NOMENCLATURE DES DIPLOMES POUR LA CODIFICATION DE L'ENQUETE EMPLOI (version au 01/01/2013)

Nomenclature de niveau détaillé en **104 postes** utilisée pour la codification des diplômes dans l'enquête emploi. Les diplômes qui ne sont plus délivrés actuellement sont en italique souligné

#### 1 Diplômes de niveau Bac+5 ou plus

- 1000 DU NIVEAU BAC+5 (DIPLOME UNIVERSITAIRE DE NIVEAU BAC+5) 1 100  
MAGISTERE
- 1110 MASTERE SPECIALISE
- 1140 DRT (DIPLOME DE RECHERCHE ET TECHNOLOGIE)
- 1200 *DEA (DIPLOME D'ETUDES APPROFONDIES), DES (DIPLOME D'ETUDES SPECIALISEE)*
- 1210 MASTER RECHERCHE (LMD)
- 1300 *DESS (DIPLOME D'ETUDES SUPERIEURES SPECIALISEES)*
- 1310 MASTER PROFESSIONNEL (LMD)
- 1420 DNSEP<sup>14</sup> NIVEAU BAC+5 (DIPLOME NATIONAL SUPERIEUR D'EXPRESSION PLASTIQUE)
- 1450 BEES 3EME DEGRE (BREVET D'ETAT D'EDUCATEUR SPORTIF 3EME DEGRE)
- 1500 CAPME (DIPLOME DE CAPACITE DE MEDECINE)
- 1600 ECOLE SUPERIEURE DE COMMERCE<sup>15</sup> NIVEAU BAC+5
- 1640 AUTRE TITRE OU CERTIFICAT NIVEAU BAC+5 (ARCHITECTE, EXPERT- COMPTABLE, DSCG, ...)
- 1700 INGENIEUR (ECOLE D'INGENIEUR)
- 1800 PROFESSEUR D'ENSEIGNEMENT SECONDAIRE<sup>16</sup> (CAPES, CAPET, CAPLP, CAFEP, PROFESSORAT DE SPORT,...) NIVEAU BAC+5
- 1880 PROFESSEUR DES ECOLES<sup>17</sup> (CAPE) NIVEAU BAC+5
- 1900 AGREGATION<sup>18</sup>
- 1960 DOCTORATS PROFESSIONS DE SANTE (MEDECINE, PHARMACIE, DENTAIRE, VETERINAIRE)
- 1970 DOCTORATS DE RECHERCHE (HORS SANTE)
- 1980 AUTRE DIPLOME NIVEAU BAC+5 OU PLUS (AVOCAT, NOTAIRE, MAGISTRAT, EXPERT GEOMETRE, SCIENCES PO<sup>19</sup>...)

#### 2 Diplômes de niveau Bac+3 ou 4 (licence, maîtrise)

- 2000 LICENCE, LICENCE GENERALE LMD (L3)
- 2010 MAITRISE, MAITRISE INTERMEDIAIRE (M1)
- 2200 *MST MAITRISE DE SCIENCES ET TECHNIQUES*
- 2300 DIPLOME D'INGENIEUR MAITRE (MAITRISE D'IUP)
- 2400 DSAA (DIPLOME SUPERIEUR DES ARTS APPLIQUES)

---

<sup>14</sup> Le DNSEP obtenu avant 2006 est classé en 2420

<sup>15</sup> Les diplômes d'écoles supérieures de commerce obtenus avant 2000 sont classés en 2600 et en 1600 à partir de 2000

<sup>16</sup> Le CAPES, CAPET, CAPLP, CAFEP, CAPEPS obtenu avant 2010 est classé en 2800 (niveau licence) et en 1800 (niveau master) à partir de 2010

<sup>17</sup> Le CAPE obtenu avant 2010 est classé en 2980 (niveau licence) et en 1880 (niveau master) à partir de 2010

<sup>18</sup> L'agrégation obtenue avant 2010 est classée en 2900 (niveau licence) et en 1900 (niveau master) à partir de 2010

<sup>19</sup> Les diplômes de sciences-po obtenus avant 2005 sont classés en 2980

- 2410 DNAT<sup>20</sup> (DIPLOME NATIONAL D'ART ET DE TECHNOLOGIE), DNAP<sup>21</sup> (DIPLOME NATIONAL D'ARTS PLASTIQUES) NIVEAU BAC+3
- 2420 DNSEP<sup>22</sup> (DIPLOME NATIONAL SUPERIEUR D'EXPRESSION PLASTIQUE) NIVEAU BAC+4
- 2450 BEES NIVEAU BAC+4 (BREVET D'ETAT D'EDUCATEUR SPORTIF 2EME DEGRE)
- 2460 DESE, DEST (DIPLOME D'ETUDES SUPERIEURES DU CNAM)
- 2500 LICENCE PRO (LICENCE PROFESSIONNELLE)
- 2560 DU NIVEAU BAC+3/4 (DIPLOME UNIVERSITAIRE DE NIVEAU BAC+3/4)
- 2580 DREA (DIPLOME DE RECHERCHE ET D'ETUDES APPLIQUEES)
- 2600 ECOLE SUPERIEURE DE COMMERCE<sup>23</sup> NIVEAU BAC+4
- 2640 AUTRE TITRE OU CERTIFICAT NIVEAU BAC+3/4 (DCG, DESCF, DECE, BEAUX- ARTS...)
- 2800 IUFM, CAPES, CAPET, AUTRES CONCOURS D'ENSEIGNEMENT SECONDAIRE (PLP2, PROFS DE SPORT...) NIVEAU BAC+3/<sup>24</sup>
- 2880 IUFM, CAPE<sup>25</sup>(PROFESSEUR DES ECOLES), ENSEIGNEMENT DU 1ER DEGRE NIVEAU BAC+3/4
- 2900 AGREGATION<sup>26</sup> NIVEAU BAC+4
- 2960 DIPLOMES DE LA SANTE ET DU TRAVAIL SOCIAL DE NIVEAU BAC+3/4 (SAGE- FEMME<sup>27</sup>, INFIRMIERE<sup>28</sup>, OSTEOPATHE, DSTS DIPLOME SUPERIEUR DE TRAVAIL SOCIAL,...)
- 2980 AUTRE DIPLOME NIVEAU BAC+3/4 (JOURNALISTE, ART, ETUDES JUDICIAIRES,...)

### 3 Diplômes de niveau Bac+2

- 3200 BTS (BREVET DE TECHNICIEN SUPERIEUR)
- 3210 DMA (DIPLOME DES METIERS D'ART)
- 3220 DTS (DIPLOME DE TECHNICIEN SUPERIEUR), DNTS, DPECF
- 3230 BTSA (BREVET DE TECHNICIEN SUPERIEUR AGRICOLE)
- 3350 BM NIVEAU BAC+2 (BREVET DE MAITRISE SUPERIEUR)
- 3400 DNAT<sup>29</sup> (DIPLOME NATIONAL D'ART ET DE TECHNOLOGIE) NIVEAU BAC+2
- 3420 DNA<sup>30</sup> (DIPLOME NATIONAL D'ARTS PLASTIQUES) NIVEAU BAC+2
- 3460 DPCT, DPCE (DIPLOMES DU CNAM NIVEAU BAC+2)
- 3500 DUT (DIPLOME UNIVERSITAIRE DE TECHNOLOGIE)
- 3510 PROPEDEUTIQUE
- 3520 DEUG (DIPLOME D'ETUDES UNIVERSITAIRES GENERALES), AUTRES DIPLOMES UNIVERSITAIRES NIVEAU BAC+2 (PCEM, DUEL, DUES,...)
- 3550 DEUST (DIPLOME D'ETUDES UNIVERSITAIRES SCIENTIFIQUES ET TECHNIQUES)
- 3560 DU NIVEAU BAC+2 (DIPLOME UNIVERSITAIRE DE NIVEAU BAC+2)
- 3630 CSA (CERTIFICAT DE SPECIALISATION AGRICOLE) NIVEAU BAC+2
- 3640 AUTRE TITRE OU CERTIFICAT NIVEAU BAC+2 (ECOLE DE VENTE, DEFA D'ARCHITECTE, CLERC DE NOTAIRE...)

<sup>20</sup> Le DNAT obtenu avant 2006 est classé en 3400

<sup>21</sup> Le DNAP obtenu avant 2006 est classé en 3420

<sup>22</sup> Le DNSEP obtenu à partir de 2006 est classé en 1420

<sup>23</sup> A partir de 2013, les diplômes d'écoles supérieures de commerce obtenus avant 2000 sont classés en 2600 et en 1600 à partir de 2000 ((codés 2600 jusqu'en 2012)

<sup>24</sup> Le CAPES, CAPET, CAPLP, CAFEP, CAPEPS obtenu avant 2010 est classé en 2800 (niveau licence), et en 1800 (niveau master) à partir de 2010

<sup>25</sup> A partir de 2013, le CAPE obtenu avant 2010 est codé en 2880 (codé 2980 jusqu'en 2012) et en 1880 (niveau master) si Obtenu à partir de 2010

<sup>26</sup> L'agrégation obtenue avant 2010 est classée en 2900 (niveau licence) et en 1900 (niveau master) à partir de 2010

<sup>27</sup> Les diplômes de sage-femme obtenus avant 2008 sont classés en 3960

<sup>28</sup> Les diplômes d'infirmière obtenus avant 2010 sont classés en 3960

<sup>29</sup> Le DNAT obtenu à partir de 2006 est classé en 2410

<sup>30</sup> Le DNAP obtenu à partir de 2006 est classé en 2410

- 3880 ECOLE NORMALE D'INSTITUTEUR<sup>31</sup>, PEGC  
3960 DIPLOMES DE LA SANTE ET DU TRAVAIL SOCIAL DE NIVEAU BAC+2 (KINE, LABORANTIN, PUERICULTRICE, EDUCATEUR, INFIRMIERE<sup>32</sup> AVANT 2010,...)  
3980 AUTRE DIPLOME NIVEAU BAC+2 (ECOLE D'ART,...)

4 Diplômes de niveau Bac

- 4000 BAC PRO (BACCALAUREAT PROFESSIONNEL)  
4010 BMA (BREVET DES METIERS D'ART)  
4020 BTM (BREVET TECHNIQUE DES METIERS)  
4030 BAC PRO AGRICOLE (BACCALAUREAT PROFESSIONNEL AGRICOLE)  
4060 BMS BREVET DES METIERS DU SPECTACLE  
4100 BEI BREVET D'ENSEIGNEMENT INDUSTRIEL  
4110 BEC BREVET D'ENSEIGNEMENT COMMERCIAL  
4120 BEH BREVET D'ENSEIGNEMENT HOTELIER  
4130 BEA BREVET D'ENSEIGNEMENT AGRICOLE  
4140 BES BREVET D'ENSEIGNEMENT SOCIAL  
4200 BT (BREVET DE TECHNICIEN)  
4230 BTA (BREVET DE TECHNICIEN AGRICOLE)  
4300 BAC TECHNO (BACCALAUREAT TECHNOLOGIQUE) : STG, STI, STL, ST2S, SMS, STT, F, G, H  
4330 BAC TECHNO AGRICOLE: STAV, STAE, STPA  
4350 BM (BREVET DE MAITRISE)  
4370 MC POST BAC (MENTION COMPLEMENTAIRE NIVEAU BAC)  
4450 BEES NIVEAU BAC (BREVET D'ETAT D'EDUCATEUR SPORTIF 1ER DEGRE), BPJEPS  
4500 BP (BREVET PROFESSIONNEL)  
4530 BPA (BREVET PROFESSIONNEL AGRICOLE)  
4600 BSEC (BREVET SUPERIEUR D'ENSEIGNEMENT COMMERCIAL)  
4630 CSA NIVEAU BAC (CERTIFICAT DE SPECIALISATION AGRICOLE NIVEAU BAC)  
4640 AUTRE TITRE OU CERTIFICAT NIVEAU BAC (BEPECASER, SECRETAIRE MEDICALE,...)  
4700 BAC GENERAL: L, ES, S, A, B, C, D, E, PHILO, MATH-ELEM, SCIENCES EX, MATHEMATIQUES ET TECHNIQUES  
4880 CAPACITE EN DROIT, DAEU (DIPLOME D'ACCES AUX ETUDES UNIVERSITAIRES), ESEU  
4900 BREVET SUPERIEUR  
4960 DIPLOMES DE LA SANTE ET DU TRAVAIL SOCIAL NIVEAU BAC (ASSISTANTE DENTAIRE, MONITEUR EDUCATEUR,...)  
4980 AUTRE DIPLOME NIVEAU BAC (PERSONNEL NAVIGANT COMMERCIAL, DIPLÔME D'ETUDES MUSICALES,...)

---

<sup>31</sup> Nouveau code pour les instituteurs a partir de 2013 (codés 3980 jusqu'en 2012)

<sup>32</sup> Les diplômes d'infirmière obtenus à partir de 2010 sont classés en 2960

5 Diplômes de niveau CAP-BEP

- 5000 CAP (CERTIFICAT D'APTITUDE PROFESSIONNELLE)
- 5030 CAPA (CERTIFICAT D'APTITUDE PROFESSIONNELLE AGRICOLE)
- 5100 BEP (BREVET D'ETUDES PROFESSIONNELLES)
- 5130 BEPA (BREVET D'ETUDES PROFESSIONNELLES AGRICOLES)
- 5230 BAA (BREVET D'APPRENTISSAGE AGRICOLE)
- 5320 BC (BREVET DE COMPAGNON)
- 5370 MC AU CAP-BEP (MENTION COMPLEMENTAIRE NIVEAU CAP-BEP)
  - 5400 PREMIERE PARTIE BACCALAUREAT, CFES (CERTIFICAT DE FIN D'ETUDES SECONDAIRES)
- 5450 BREVET ELEMENTAIRE, BEPS BREVET ENSEIGNEMENT PRIMAIRE SUPERIEUR
- 5530 BPA (BREVET PROFESSIONNEL AGRICOLE)
- 5560 EFAA (EXAMEN DE FIN D'APPRENTISSAGE ARTISANAL)
- 5630 CSA NIVEAU CAP-BEP (CERTIFICAT DE SPECIALISATION AGRICOLE NIVEAU CAP/BEP)
- 5640 AUTRE TITRE OU CERTIFICAT NIVEAU CAP-BEP (TITRE AFPA NIVEAU CAP,...)
- 5960 DIPLOME SANTE ET TRAVAIL SOCIAL NIVEAU CAP-BEP (AIDE SOIGNANTE, AUXILIAIRE PUERICULTRICE,...)
- 5980 AUTRE DIPLÔME NIVEAU CAP-BEP (MONITEUR AUTO ECOLE<sup>33</sup>, ENSEIGNEMENT MENAGER,...)

6 Diplômes de niveau Brevet

- 6400 DNB (DIPLOME NATIONAL BREVET), BREVET DES COLLEGES, BEPC
- 6410 CEPRO (CERTIFICAT D'EDUCATION PROFESSIONNELLE)

7 Diplômes de niveau CEP ou aucun diplôme

- 7400 CEP (CERTIFICAT D'ETUDES PRIMAIRES), DFEO (DIPLOME FIN ETUDES OBLIGATOIRES)
- 7510 CFG (CERTIFICAT DE FORMATION GENERALE)
- 7990 AUCUN DIPLÔME RECONNU

---

<sup>33</sup> un nouveau titre de niveau bac a été créé en 2001 qui remplace celui de moniteur auto-école : le brevet pour l'exercice de la profession d'enseignant de la conduite automobile et de la sécurité routière (BEPECASER).

## Annexe 4 – Nomenclature des niveaux d'études

### NOMENCLATURE DES NIVEAUX D'ETUDES UTILISEE POUR LA CODIFICATION DE L'ENQUETE EMPLOI (version au 01/01/2013)

Cette nomenclature de niveaux en **137 postes** n'a pas de caractère officiel, elle est utilisée par l'Insee pour la codification des enquêtes ménages. Les niveaux des diplômes qui ne sont plus délivrés actuellement sont en italique souligné

Les années terminales de préparation des diplômes ont un code dont les 4 premiers caractères sont identiques à ceux du code diplôme, et le dernier est égal à 0

Les années non terminales de préparation des diplômes ont un code dont le 5<sup>e</sup> caractère est différent de 0. Elles sont surlignées en gras.

#### 1 Formations de niveau Bac+5 ou plus

10000 DU NIVEAU BAC+5 (DIPLOME UNIVERSITAIRE NIV BAC+5)

11000 MAGISTERE

11100 MASTERE SPECIALISE

11400 DRT (DIPLOME RECHERCHE TECHNOLOGIE)

12000 DEA (DIPLOME ETUDES APPROFONDIES), *DES (DIPLOME ETUDES SPECIALISEES)*

12100 MASTER RECHERCHE (LMD)

13000 DESS DIPLOME ETUDES SUPERIEURES SPECIALISEES

13100 MASTER PROFESSIONNEL (LMD)

14200 DNSEP<sup>34</sup> NIVEAU BAC+5 (DIPLOME NATIONAL SUPERIEUR D'EXPRESSION PLASTIQUE)

14500 BEES NIVEAU BAC+5 (BREVET ETAT EDUCATEUR SPORTIF 3EME DEGRE)

15000 CAPME (CAPACITE DE MEDECINE)

16000 ECOLE SUPERIEURE DE COMMERCE<sup>35</sup> NIVEAU BAC+5

16400 AUTRE TITRE HOMOLOGUE BAC +5 (ARCHITECTE, EXPERT-COMPTABLE, DSCG,...)

17000 INGENIEUR (ECOLE D'INGENIEUR)

18000 PROFESSEUR D'ENSEIGNEMENT SECONDAIRE<sup>36</sup> NIVEAU MASTER : CAPES, CAPET, CAPLP, CAFEP, PROFESSORAT DE SPORT...

18800 PROFESSEUR D'ENSEIGNEMENT PRIMAIRE<sup>37</sup> NIVEAU MASTER : CAPE

19000 AGREGATION<sup>38</sup>

19600 DOCTORAT DES PROFESSIONS DE SANTE (MEDECINE, PHARMACIE, DENTAIRE, VETERINAIRE)

19700 DOCTORAT DE RECHERCHE (HORS SANTE)

19800 AUTRES DIPLOMES DE NIVEAU BAC+5 OU PLUS (AVOCAT, NOTAIRE, MAGISTRAT, SCIENCES PO<sup>39</sup>,...)

**19991 ANNEE NON TERMINALE DE DOCTORAT DE RECHERCHE (HORS SANTE)**

**19992 ANNEE NON TERMINALE DE DOCTORAT MEDECINE, PHARMACIE, DENTAIRE,...**

**19999 BAC+5 ET PLUS (SANS AUTRE INDICATION)**

<sup>34</sup> Le DNSEP obtenu avant 2006 est classé en 24200 et en 14200 à partir de 2006

<sup>35</sup> Les diplômes d'écoles supérieures de commerce obtenus avant 2000 sont classés en 26000 et en 16000 à partir de 2000

<sup>36</sup> Le CAPES, CAPET, CAPLP, CAFEP, CAPEPS obtenu avant 2010 est classé en 28 000 (niveau licence) et en 18 000 (niveau master) à partir de 2010

<sup>37</sup> Le CAPE obtenu avant 2010 est classé en 29800 (niveau licence) et en 18800 (niveau master) à partir de 2010

<sup>38</sup> L'agrégation obtenue avant 2010 est classée en 29000 (niveau licence) et en 19000 (niveau master) à partir de 2010

<sup>39</sup> Les diplômes de sciences po obtenus *avant* 2005 sont classés en 29800 et en 19800 à partir de 2005

## 2 Formations de niveau Bac+3 et Bac+4

- 20000 LICENCE, LICENCE GENERALE LMD (L3)  
 20100 MAITRISE, 1ERE ANNEE MASTER (M1)  
 22000 MST (MAITRISE SCIENCES TECHNIQUES)  
 23000 DIPLOME D'INGENIEUR MAITRE (MAITRISE D IUP)  
 24000 DSAA (DIPLOME SUPERIEUR ARTS APPLIQUES)  
 24100 DNAT<sup>40</sup> (DIPLOME NATIONAL D'ART ET DE TECHNOLOGIE), DNAP<sup>41</sup>  
 (DIPLÔME NATIONAL D'ARTS PLASTIQUES) NIVEAU BAC+3  
 24200 DNSEP<sup>42</sup> (DIPLOME NATIONAL SUPERIEUR EXPRESSION PLASTIQUE) NIVEAU BAC+4  
 24500 BEES NIVEAU BAC+4 (BREVET ETAT EDUCATEUR SPORTIF 2EME DEGRE)  
 24600 DESE, DEST (DIPLOME D'ETUDES SUPERIEURES DU CNAM)  
 25000 LICENCE PRO (LICENCE PROFESSIONNELLE)  
 25600 DU NIVEAU BAC+3/4 (DIPLOME UNIVERSITAIRE NIV BAC+3/4)  
 25800 DREA (DIPLOME DE RECHERCHE ET D'ETUDES APPLIQUEES)  
 26000 ECOLE SUPERIEURE DE COMMERCE<sup>43</sup> NIVEAU BAC+4  
 26400 AUTRE TITRE HOMOLOGUE BAC+3/4 (DCG, DESCF, DECF, BEAUX-ARTS,...)  
 28000 IUFM, CAPES, CAPET, CAPLP, CAFEP, PROFESSORAT DE SPORT... NIVEAU LICENCE<sup>44</sup>  
 28800 IUFM. CAPE<sup>45</sup> (PROFESSEUR DES ECOLES), ENSEIGNEMENT PRIMAIRE NIVEAU LICENCE  
 29000 AGREGATION<sup>46</sup>  
 29600 DIPLOMES DE LA SANTE ET DU TRAVAIL SOCIAL DE NIVEAU BAC+3/4 (SAGE- FEMME<sup>47</sup>,  
 INFIRMIERE<sup>48</sup> OSTEOPATHE, DSTS DIPLOME SUPERIEUR DE TRAVAIL SOCIAL)  
 29800 AUTRES DIPLOMES DE NIVEAU BAC+3/4 (JOURNALISTE, ART, ETUDES  
 JUDICIAIRES,...)  
**29991 1ERE ANNEE IUFM**  
**29998 BAC+3 (SANS AUTRE INDICATION)**  
**29999 BAC+4 (SANS AUTRE INDICATION)**

## 3 Formations de niveau Bac+2

- 32000 BTS (BREVET DE TECHNICIEN SUPERIEUR)  
 32100 DMA (DIPLOME DES METIERS D'ART)  
 32200 DTS (DIPLOME DE TECHNICIEN SUPERIEUR), DNTS, DPECF  
 32300 BTSA (BREVET DE TECHNICIEN SUPERIEUR AGRICOLE)  
 33500 BM NIVEAU BAC+2 (BREVET DE MAITRISE SUPERIEUR)  
 34000 DNAT (DIPLOME NATIONAL ART ET TECHNOLOGIE)  
 34200 DNAP (DIPLOME NATIONAL D'ARTS PLASTIQUES)  
 34600 DPCT, DPCE (DIPLOME PREMIER CYCLE CNAM)  
 35000 DUT (DIPLOME UNIVERSITAIRE TECHNOLOGIE)

<sup>40</sup> Le DNAT obtenu avant 2006 est classé en 34000 et en 24100 à partir de 2006

<sup>41</sup> Le DNAP obtenu avant 2006 est classé en 34200 et en 24100 à partir de 2006

<sup>42</sup> Le DNSEP obtenu avant 2006 est classé en 24200 et en 14200 à partir de 2006

<sup>43</sup> Les diplômes d'écoles supérieures de commerce obtenus avant 2000 sont classés en 26000 et en 16000 à partir de 2000

<sup>44</sup> Le CAPES, CAPET, CAPLP, CAFEP, CAPEPS obtenu avant 2010 est classé en 28000 (niveau licence), et en 18000 (niveau master) à partir de 2010

<sup>45</sup> Le CAPE obtenu avant 2010 est classé en 2980 (niveau licence) et en 1880 (niveau master) à partir de 2010

<sup>46</sup> L'agrégation obtenue avant 2010 est classée en 29000 (niveau licence) et en 19000 (niveau master) à partir de 2010

<sup>47</sup> Le diplôme de sage-femme obtenu avant 2008 est classé en 39600 et en 29600 à partir de 2008

<sup>48</sup> Le diplôme d'infirmière obtenu avant 2010 est classé en 39600 et en 29600 à partir de 2010



- 35100 PROPEDEUTIQUE
- 35200 DIPLOMES UNIVERSITAIRES GENERAUX 1ER CYCLE (DEUG, DUEL, DUES, ...), 2EME ANNEE LICENCE (L2)
- 35500 DEUST (DIPLOME ETUDES UNIVERSITAIRES SCIENTIFIQUES TECHNIQUES)
- 35600 DU NIVEAU BAC+2 (DIPLOME UNIVERSITAIRE DE NIVEAU BAC+2)
- 36300 CSA NIVEAU BAC+2 (CERTIFICAT SPECIALISATION AGRICOLE NIVEAU BAC+2)
- 36400 AUTRE TITRE HOMOLOGUE DE NIVEAU BAC+2 (ECOLE DE VENTE, DEFA D'ARCHITECTE, CLERC DE NOTAIRE,...)
- 38800 ECOLE NORMALE D'INSTITUTEUR<sup>49</sup> PEGC
- 38881 ANNEE POST BTS POST DUT FCIL NIVEAU BAC+2**
- 38883 ANNEE NON TERMINALE ECOLE SUPERIEURE DE COMMERCE**
- 38884 ANNEE NON TERMINALE ECOLE INGENIEUR**
- 38885 ANNEE NON TERMINALE DES AUTRES FORMATIONS DE NIVEAU > AU BAC+2 (ARCHITECTURE, MAGISTRATURE, VETERINAIRE, SCIENCES PO, JOURNALISTE,...)**
- 39600 DIPLOMES DE LA SANTE ET DU TRAVAIL SOCIAL DE NIVEAU BAC+2 (KINE, LABORANTIN, PUERICULTRICE, EDUCATEUR, INFIRMIERE<sup>50</sup> AVANT 2010,...)
- 39800 AUTRE FORMATION NIVEAU BAC+2 (ECOLE D'ART NIVEAU BAC+2,...)
- 39991 DCEM 2EME CYCLE MEDECINE ODONTOLOGIE PHARMACIE**
- 39999 BAC+2 (SANS AUTRE INDICATION)**

4 Formations de niveau Bac et Bac+1

- 40000 BAC PRO (BACCALAUREAT PROFESSIONNEL)
- 40100 BMA (BREVET DES METIERS D'ART)
- 40200 BTM (BREVET TECHNIQUE DES METIERS)
- 40300 BAC PRO AGRICOLE (BACCALAUREAT PROFESSIONNEL AGRICOLE)
- 40600 BMS BREVET DES METIERS DU SPECTACLE
- 41000 BEI BREVET D'ENSEIGNEMENT INDUSTRIEL
- 41100 BEC BREVET D'ENSEIGNEMENT COMMERCIAL
- 41200 BEH BREVET D'ENSEIGNEMENT HOTELIER
- 41300 BEA BREVET D'ENSEIGNEMENT AGRICOLE
- 41400 BES BREVET D'ENSEIGNEMENT SOCIAL
- 42000 BT (BREVET DE TECHNICIEN)
- 42300 BTA (BREVET DE TECHNICIEN AGRICOLE)
- 43000 BAC TECHNO (BACCALAUREAT TECHNOLOGIQUE) : STG, STI, STL, ST2S, SMS, STT, F, G, H
- 43300 BAC TECHNO AGRICOLE: STAV, STAE, STPA
- 43500 BM (BREVET DE MAITRISE)
- 43700 MC APRES BAC (MENTION COMPLEMENTAIRE NIVEAU BAC)
- 44500 BEES NIVEAU BAC (BREVET ETAT EDUCATEUR SPORTIF PREMIER DEGRE), BPJEPS
- 45000 BP (BREVET PROFESSIONNEL)
- 45300 BPA (BREVET PROFESSIONNEL AGRICOLE)
- 46000 BSEC (BREVET SUPERIEUR ENSEIGNEMENT COMMERCIAL)
- 46300 CSA NIVEAU BAC (CERTIFICAT DE SPECIALISATION AGRICOLE NIVEAU BAC)
- 46400 AUTRE TITRE HOMOLOGUE NIVEAU BAC (BEPECASER, SECRETAIRE MEDICALE,...)
- 47000 BAC GENERAL: L, ES, S, A, B, C, D, E, PHILO, MATH-ELEM, SCIENCES EX, BREVET SUPERIEUR

<sup>49</sup> Nouveau code pour les instituteurs à partir de 2013 (codés 39800 jusqu'en 2012)

<sup>50</sup> Le diplôme d'infirmière obtenu avant 2010 est classé en 39600 et en 29600 à partir de 2010

**48201 ANNEE NON TERMINALE DEUST, DUT**  
**48202 ANNEE NON TERMINALE BTS, BTSA**  
**48203 ANNEE NON TERMINALE DIPLOMES DU TRAVAIL SOCIAL OU DE LA SANTE DE NIVEAU BAC+2 OU BAC+3/4**  
**48204 ANNEE NON TERMINALE AUTRE CYCLE NIVEAU BAC+2 (ART, INSTITUTEUR,..)**  
48800 CAPACITE EN DROIT, DAEU, ESEU  
49600 DIPLOMES DE LA SANTE ET DU TRAVAIL SOCIAL NIVEAU BAC (ASSISTANTE DENTAIRE, MONITEUR EDUCATEUR,...)  
49800 AUTRE DIPLOME NIVEAU BAC (PERSONNEL NAVIGANT COMMERCIAL, DIPLÔME D'ETUDES MUSICALES,...)  
**49991 CPGE (CLASSE PREPARATOIRE GRANDES ECOLES)**  
**49992 1ERE ANNEE DEUG, 1ERE ANNEE LICENCE (L1)**  
**49993 FORMATION COMPLEMENTAIRE POST-BAC : CLASSE PREPARATOIRE AUX ECOLES DE SANTE, AUX CONCOURS,..**  
**49999 BAC+1 (SANS AUTRE INDICATION)**

5 Formations de niveau CAP, BEP, seconde et première de lycée

50000 CAP (CERTIFICAT D'APTITUDE PROFESSIONNELLE)  
50300 CAPA (CERTIFICAT D'APTITUDE PROFESSIONNELLE AGRICOLE)  
51000 BEP (BREVET D'ETUDES PROFESSIONNELLES)  
51300 BEPA (BREVET D'ETUDES PROFESSIONNELLES AGRICOLES)  
52300 BAA (BREVET APPRENTISSAGE AGRICOLE)  
53200 BC (BREVET COMPAGNON)  
53700 MC AU CAP, BEP (MENTION COMPLEMENTAIRE NIVEAU CAP-BEP)  
54500 BE (BREVET ELEMENTAIRE), BEPS (BREVET ENSEIGNEMENT PRIMAIRE SUPERIEUR)  
55300 BPA (BREVET PROFESSIONNEL AGRICOLE)  
55600 EFAA (EXAMEN FIN APPRENTISSAGE ARTISANAL)  
56300 CSA NIVEAU CAP-BEP (CERTIFICAT DE SPECIALISATION AGRICOLE NIVEAU CAP-BEP)  
56400 AUTRE TITRE HOMOLOGUE NIVEAU CAP-BEP (TITRE AFPA NIVEAU CAP,...)  
**58881 SECONDE TECHNOLOGIQUE**  
**58882 SECONDE PRO<sup>51</sup> (SECONDE PROFESSIONNELLE), SECONDE BT, SECONDE SPECIALE**  
**58883 1ERE TECHNO, PREMIERE ADAPTATION**  
**58884 PREMIERE BAC PRO, PREMIERE BT**  
**58885 ANNEE NON TERMINALE TITRE PROFESSIONNEL NIVEAU BAC (BP,...)**  
**58886 FORMATION COMPLEMENTAIRE POST CAP-BEP**  
59600 DIPLOME SANTE ET TRAVAIL SOCIAL NIV CAP-BEP (AIDE SOIGNANTE, AUXILIAIRE PUERICULTRICE,...)  
59800 AUTRE FORMATION NIVEAU CAP-BEP (MONITEUR AUTO ECOLE<sup>52</sup>, ENSEIGNEMENT MENAGER,...)  
**59991 SECONDE GENERALE, SECONDE INDETERMINEE**  
**59992 1ERE GENERALE**

---

<sup>51</sup> La seconde professionnelle obtenue avant 2010 est codée en 68882 et en 58882 à partir de 2010

<sup>52</sup> un nouveau titre de niveau bac a été créé en 2001 qui remplace celui de moniteur auto-école: le brevet pour l'exercice de la profession d'enseignant de la conduite automobile et de la sécurité routière (BEPECASER) classé en 46400

6 Formation de niveau Brevet des collèges et année non terminale de CAP ou BEP

64000 TROISIEME, DNB, *BEPC, BREVET DES COLLEGES*

64100 *CEPRO (CERTIFICAT EDUCATION PROFESSIONNELLE)*

**68881 ANNEE NON TERMINALE CAP**

**68882 ANNEE NON TERMINALE BEP**

**68883 ANNEE NON TERMINALE AUTRE DIPLOME NIVEAU CAP-BEP**

7 Formation de niveau CEP, 6ème, 5ème, ou 4ème de collège, ou aucun diplôme.

74000 *CEP (CERTIFICAT D'ETUDES PRIMAIRES), DFEO (DIPLOME FIN ETUDES OBLIGATOIRES)*

**78882 CPPN, CPA, CLASSE TRANSITION**

79900 AUCUN DIPLÔME RECONNU

**79991 ETUDES PRIMAIRES**

**79992 6EME, 5EME**

**79993 4EME**

**79995 IME (INSTITUT MEDICO EDUCATIF), IMP (INSTITUT MEDICO PEDAGOGIQUE), IMPRO (INSTITUT MEDICO PROFESSIONNEL)**

## Annexe 5 – Nomenclature des spécialités

### Nomenclature des groupes de Spécialités de Formation (NSF : niveau 100)

#### **1 DOMAINES DISCIPLINAIRES**

##### **10 Formations générales**

100 Formations générales

##### **11 Mathématiques et sciences**

110 Spécialités pluri-scientifiques

111 Physique-Chimie

112 Chimie-biologie, biochimie

113 Sciences naturelles (biologie-géologie)

114 Mathématiques

115 Physique

116 Chimie

117 Sciences de la terre

118 Sciences de la vie

##### **12 Sciences humaines et droit**

120 Spécialités pluridisciplinaires sciences humaines et droit

121 Géographie

122 Economie

123 Sciences sociales (y compris démographie, anthropologie)

124 Psychologie

125 Linguistique

126 Histoire

127 Philosophie, éthique et théologie

128 Droit, sciences politiques

##### **13 Lettres et arts**

130 Spécialités littéraires et artistiques plurivalentes

131 Français, littérature et civilisation française

132 Arts plastiques

133 Musique, arts du spectacle

134 Autres disciplines artistiques et spécialités artistiques plurivalentes

135 Langues et civilisations anciennes

136 Langues vivantes, civilisations étrangères et régionales

#### **2 DOMAINES TECHNICO-PROFESSIONNELS DE LA PRODUCTION**

##### **20 Spécialités pluri-technologiques de la production**

200 Technologies industrielles fondamentales (génie industriel et procédés de transformation, spécialités à dominante fonctionnelle)

201 Technologies de commandes des transformations industrielles (automatismes et robotique industriels, informatique industrielle)

##### **21 Agriculture, pêche, forêts et espaces verts**

210 Spécialités plurivalentes de l'agronomie et de l'agriculture

211 Productions végétales, cultures spécialisées et protection des cultures (horticulture, viticulture, arboriculture fruitière...)

212 Production animales, élevage spécialisé, aquaculture, soins aux animaux yc vétérinaire

213 Forêts, espaces naturels, faune sauvage, pêche

214 Aménagement paysager (parcs, jardins, espaces verts, terrains de sport)

##### **22 Transformations**

220 Spécialités pluritechnologiques des transformations

221 Agro-alimentaire, alimentation, cuisine.

222 Transformations chimiques et apparentées (y compris industrie pharmaceutique)

- 223 Métallurgie (y compris sidérurgie, fonderie, non-ferreux..)
- 224 Matériaux de construction, verre, céramique
- 225 Plasturgie, matériaux composites
- 226 Papier, carton
- 227 Energie, génie climatique (y compris énergie nucléaire, thermique, hydraulique, utilités : froid, climatisation, chauffage)

### **23 Génie civil, construction et bois**

- 230 Spé. pluritechnologiques, génie civil, construction, bois
- 231 Mines et carrières, génie civil, topographie
- 232 Bâtiment : construction et couverture
- 233 Bâtiment : finitions
- 234 Travail du bois et de l'ameublement

### **24 Matériaux souples**

- 240 Spécialités pluritechnologiques matériaux souples
- 241 Textile
- 242 Habillement (y compris mode, couture)
- 243 Cuirs et peaux

### **25 Mécanique, électricité, électronique**

- 250 Spé pluritechnologiques mécanique-électricité (y compris maintenance mécano-électrique)
- 251 Mécanique générale et de précision, usinage
- 252 Moteurs et mécanique auto
- 253 Mécanique aéronautique et spatiale
- 254 Structures métalliques (y compris soudure, carrosserie, coque bateau, cellule avion)
- 255 Electricité, électronique (non compris automatismes, productique)

## **3 DOMAINES TECHNICO-PROFESSIONNELS DES SERVICES**

### **30 Spécialités plurivalentes des services**

- 300 Spécialités plurivalentes des services

### **31 Echanges et gestion**

- 310 Spécialités plurivalentes des échanges et de la gestion (y.c administration générale des entreprises et des collectivités)
- 311 Transport, manutention, magasinage
- 312 Commerce, vente
- 313 Finances, banque, assurances
- 314 Comptabilité, gestion
- 315 Ressources humaines, gestion du personnel, gestion de l'emploi

### **32 Communication et information**

- 320 Spécialités plurivalentes de la communication
- 321 Journalisme et communication (y compris communication graphique et publicité)
- 322 Techniques de l'imprimerie et de l'édition
- 323 Techniques de l'image et du son, métiers connexes du spectacle
- 324 Secrétariat, bureautique
- 325 Documentation, bibliothèques, administration des données
- 326 Informatique, traitement de l'information, réseaux de transmission des données

### **33 Services aux personnes**

- 330 Spécialités plurivalentes sanitaires et sociales
- 331 Santé
- 332 Travail social
- 333 Enseignement, formation
- 334 Accueil, hôtellerie, tourisme
- 335 Animation culturelle, sportive et de loisirs
- 336 Coiffure, esthétique et autres spécialités des services aux personnes

**34 Services à la collectivité**

- 340 Spécialités plurivalentes des services à la collectivité
- 341 Aménagement du territoire, développement, urbanisme
- 342 Protection et développement du patrimoine
- 343 Nettoyage, assainissement, protection de l'environnement
- 344 Sécurité des biens et des personnes, police, surveillance (y compris hygiène et sécurité)
- 345 Application des droits et statuts des personnes
- 346 Spécialités militaires

**4 DOMAINES DU DÉVELOPPEMENT PERSONNEL**

**41 Domaines des capacités individuelles**

- 410 Spécialités concernant plusieurs capacités
- 411 Pratiques sportives (y compris arts martiaux)
- 412 Développement des capacités mentales et apprentissages de base
- 413 Développement des capacités comportementales et relationnelles
- 414 Développement des capacités individuelles d'organisation
- 415 Développement des capacités d'orientation, d'insertion ou de réinsertion sociales et professionnelles

**42 Domaines des activités quotidiennes et de loisirs**

- 421 Jeux et activités spécifiques de loisirs
- 422 Economie et activités domestiques
- 423 Vie familiale, vie sociale et autres formations au développement personnel

**Annexe 6 – Nomenclature NAF rev2 en 88 divisions (A88)**

Section	Libellé des sections	Code Division	Intitulé
A	AGRICULTURE, SYLVICULTURE ET PÊCHE	01	Culture et production animale, chasse et services annexes
		02	Sylviculture et exploitation forestière
		03	Pêche et aquaculture
B	INDUSTRIES EXTRACTIVES	05	Extraction de houille et de lignite
		06	Extraction d'hydrocarbures
		07	Extraction de minerais métalliques
		08	Autres industries extractives
		09	Services de soutien aux industries extractives
C	INDUSTRIE MANUFACTURIÈRE	10	Industries alimentaires
		11	Fabrication de boissons
		12	Fabrication de produits à base de tabac
		13	Fabrication de textiles
		14	Industrie de l'habillement
		15	Industrie du cuir et de la chaussure
		16	Travail du bois et fabrication d'articles en bois et en liège, à l'exception des meubles ; fabrication d'articles en vannerie et sparterie
		17	Industrie du papier et du carton
		18	Imprimerie et reproduction d'enregistrements
		19	Cokéfaction et raffinage
		20	Industrie chimique
		21	Industrie pharmaceutique
		22	Fabrication de produits en caoutchouc et en plastique
		23	Fabrication d'autres produits minéraux non métalliques
		24	Métallurgie
		25	Fabrication de produits métalliques, à l'exception des machines et des équipements
		26	Fabrication de produits informatiques, électroniques et optiques
		27	Fabrication d'équipements électriques
		28	Fabrication de machines et équipements n.c.a.
29	Industrie automobile		
30	Fabrication d'autres matériels de transport		
31	Fabrication de meubles		
32	Autres industries manufacturières		
33	Réparation et installation de machines et d'équipements		
D	PRODUCTION ET DISTRIBUTION D'ÉLECTRICITÉ, DE GAZ, DE VAPEUR ET D'AIR CONDITIONNÉ	35	Production et distribution d'électricité, de gaz, de vapeur et d'air conditionné
E	PRODUCTION ET DISTRIBUTION D'EAU ; ASSAINISSEMENT, GESTION DES DÉCHETS ET DÉPOLLUTION	36	Captage, traitement et distribution d'eau
		37	Collecte et traitement des eaux usées
		38	Collecte, traitement et élimination des déchets ; récupération
		39	Dépollution et autres services de gestion des déchets
F	CONSTRUCTION	41	Construction de bâtiments
		42	Génie civil
		43	Travaux de construction spécialisés
G	COMMERCE ; RÉPARATION D'AUTOMOBILES ET DE MOTOCYCLES	45	Commerce et réparation d'automobiles et de motocycles
		46	Commerce de gros, à l'exception des automobiles et des motocycles
		47	Commerce de détail, à l'exception des automobiles et des motocycles

H	TRANSPORTS ET ENTREPOSAGE	49 50 51 52 53	Transports terrestres et transport par conduites Transports par eau Transports aériens Entreposage et services auxiliaires des transports Activités de poste et de courrier
I	HÉBERGEMENT ET RESTAURATION	55 56	Hébergement Restauration
J	INFORMATION ET COMMUNICATION	58 59 60 61 62 63	Édition Production de films cinématographiques, de vidéo et de programmes de télévision ; enregistrement sonore et édition musicale Programmation et diffusion Télécommunications Programmation, conseil et autres activités informatiques Services d'information
K	ACTIVITÉS FINANCIÈRES ET D'ASSURANCE	64 65 66	Activités des services financiers, hors assurance et caisses de retraite Assurance Activités auxiliaires de services financiers et d'assurance
L	ACTIVITÉS IMMOBILIÈRES	68	Activités immobilières
M	ACTIVITÉS SPÉCIALISÉES, SCIENTIFIQUES ET TECHNIQUES	69 70 71 72 73 74 75	Activités juridiques et comptables Activités des sièges sociaux ; conseil de gestion Activités d'architecture et d'ingénierie ; activités de contrôle et analyses techniques Recherche-développement scientifique Publicité et études de marché Autres activités spécialisées, scientifiques et techniques Activités vétérinaires
N	ACTIVITÉS DE SERVICES ADMINISTRATIFS ET DE SOUTIEN	77 78 79 80 81 82	Activités de location et location-bail Activités liées à l'emploi Activités des agences de voyage, voyagistes, services de réservation et activités connexes Enquêtes et sécurité Services relatifs aux bâtiments et aménagement paysager Activités administratives et autres activités de soutien aux entreprises
O	ADMINISTRATION PUBLIQUE	84	Administration publique et défense ; sécurité sociale obligatoire
P	ENSEIGNEMENT	85	Enseignement
Q	SANTÉ HUMAINE ET ACTION SOCIALE	86 87 88	Activités pour la santé humaine Hébergement médico-social et social Action sociale sans hébergement
R	ARTS, SPECTACLES ET ACTIVITÉS RÉCRÉATIVES	90 91 92 93	Activités créatives, artistiques et de spectacle Bibliothèques, archives, musées et autres activités culturelles Organisation de jeux de hasard et d'argent Activités sportives, récréatives et de loisirs
S	AUTRES ACTIVITÉS DE SERVICES	94 95 96	Activités des organisations associatives Réparation d'ordinateurs et de biens personnels et domestiques Autres services personnels
T	ACTIVITÉS DES MÉNAGES EN TANT QU'EMPLOYEURS ; ACTIVITÉS INDIFFÉRENCIÉES DES MÉNAGES EN TANT QUE PRODUCTEURS DE BIENS ET SERVICES POUR USAGE PROPRE	97 98	Activités des ménages en tant qu'employeurs de personnel domestique Activités indifférenciées des ménages en tant que producteurs de biens et services pour usage propre
U	ACTIVITÉS EXTRA-TERRITORIALES	99	Activités des organisations et organismes extraterritoriaux



## ANNEXE 7 – Procédure de codification de l'activité selon la NAF rev2

### SICORE ACTIVITE

(document de travail)

Procédure de codification des séquences de la Génération 2010 à 3 ans

#### **A- Définition du fichier :**

Ensemble des séquences en entreprise (29 237 séquences concernées) :

- Entreprise non trouvée dans le menu entreprise
- Réponses aux questions EP3 à EP7 **avec précision** de l'activité en clair en ep10

#### **B- Traitement du fichier en entrée :**

##### **1) Définition des catégories à partir des réponses aux questions EP3 à EP7**

- 01 Public : entreprise nationalisée
- 02 Public : autre administration
- 03 Fabrication pdt
- 04 Vente service entreprise
- 05 Vente service autre
- 06 Vente service particulier
- 07 Commerce pdt entreprise
- 08 Commerce pdt indéterminé
- 09 Commerce pdt non alim particulier
- 10 Commerce pdt indéterminé particulier
- 11 Commerce pdt alim particulier lieu indéterminé
- 12 Commerce pdt alim & non alim particulier lieu indéterminé
- 13 Autres activités

Un regroupement a été nécessaire pour définir des mots clés pour préciser l'activité en clair pour le codage dans Sicore. Pour certaines catégories, les mots clés non pas été retenus car non pertinent dans le codage de Sicore.

#### ***Mots clés utilisés :***

- 1- FABRICATION : catégorie 03
- 2- COMMERCE GROS : catégorie 07 (commerce interentreprises)
- 3- COMMERCE : catégorie 08
- 4- COMMERCE DETAIL : catégories 09 10 11 et 12 (commerce pour des particuliers)

Sicore utilise des synonymes de ces termes :

- 1- PRODUCTION
- 2- DISTRIBUTION
- 3- COMMERCE
- 4- DETAIL

## 2) Correction orthographique et suppression des accents et caractères spéciaux sur le libellé en clair en ep10

Procédure effectuée à partir d'Excel avec l'utilisation du correcteur orthographique.

### C- Sicore

Utilisation de l'environnement Sicore activité « APE\_2012\_06.ENV »  
(passer par environnement -> TEST puis charger l'environnement)

Pour coder l'activité, Sicore n'utilise qu'une seule variable LIBELLE. C'est pourquoi plusieurs fichiers en entrée ont été testés pour coder un maximum de séquences.

- Fichier 1 :  
LIBELLE : libellé en clair (brut) de l'activité (sans normalisation de Sicore)

Pour les fichiers suivants, la normalisation de Sicore a permis de définir les 6 mots déterminants de chaque libellé brut (en supprimant notamment les mots de liaison).

A partir du fichier 1 en sortie, le libellé en clair est construit en concaténant les mots déterminants.

- Fichier 2 :  
LIBELLE : mot clé de la catégorie + libellé en clair de l'activité (reconstruit) OU libellé en clair de l'activité (pour les catégories sans mots clés)
- Fichier 3 :  
LIBELLE : 2 premiers mots du fichier 2
- Fichier 4 :  
LIBELLE : premier mot du fichier 2 à l'exception du mot clé

### D- Codification selon la nomenclature d'activité NAF rev2 (88 divisions)

L'ensemble des séquences ont été codées 4 fois. Dans chaque fichier codé, Sicore fournit entre 0 et 5 échos.

#### 1) Choix du bon code APE dans chacun des 4 fichiers

Résultats des 4 fichiers codés par Sicore :

Fichier 1 :

- Codés : 57,68% (16 863)
- Erreur de redondance : 10,37% (3 033)
- Echec de codage : 31,95% (9341)

Fichier 2 :

- Codés : 53,30% (15 583)
- Erreur de redondance : 11,28% (3 298)
- Echec de codage : 35,42% (10 356)

Fichier 3 :

- Codés : 61,31% (17 926)

- Erreur de redondance : 8,28% (2 420)
- Echec de codage : 30,41% (8 891)

Fichier 4 :

- Codés : 76,97% (22 505)
- Erreur de redondance : 6,45% (1 887)
- Echec de codage : 16,57% (4 845)

Au vu des résultats, 2 règles ont été définies et appliquées dans les 4 fichiers pour choisir le meilleur écho :

- 1ère règle : si 1 SEUL écho dont la probabilité est supérieure à 40, l'écho en 5 positions est choisi.
- 2ème règle : si échos multiples :
  - o sélection des 2 premières positions des échos
  - o somme des probabilités des échos identiques SUCCESSIFS (premier écho inclus)

Si la somme est supérieure à 40, l'écho en 2 positions est

choisi. Au final, pour chaque séquence :

- un code APE en 5 positions
- un code APE en 2 positions
- pas de code

## 2) Reprise de codes APE sur le fichier 2 et 3

Après analyse des résultats, les règles ont été élargies afin de récupérer des codes du fichier 2 et 3 pour les séquences qui n'ont pas été codées.

- 1ère règle : si le premier écho dont la probabilité est supérieure à 40, l'écho en 5 positions est choisi.
- 2ème règle : si échos multiples :
  - o sélection des 2 premières positions des échos
  - o comparaison des différents échos (1 à 5)
  - o somme des probabilités des échos identiques NON SUCCESSIFS (premier écho inclus)

Si la somme est supérieure à 40, l'écho en 2 positions est choisi.

## 3) Choix du meilleur code APE

A l'issue des étapes précédentes, un fichier global contenant les 4 résultats de codage a été construit.

- Le code du fichier 2 a été choisi en priorité 1
- Le code du fichier 3 a été choisi en priorité 2
- Le code du fichier 4 a été choisi si code absent dans le fichier 1 et inversement
- Si le code du fichier 4 et celui du fichier 1 sont égaux le code a été choisi (sur 2 ou 5 positions)

- Si le code du fichier 4 et celui du fichier 1 sont différents aucun code n'a été choisi (85 séquences concernées)

Une étape d'imputation s'est avérée nécessaire pour les catégories suivantes (1 seul code sur 2 positions selon la nomenclature NAF rev2) :

- commerce gros ou détail pour les automobiles et motocycles : 45
- commerce gros : 46
- commerce détail : 47

Lorsque les séquences des 2 dernières catégories ne sont pas codées alors on impute le code APE correspondant (46 ou 47), excepté celles dont l'activité en clair contient les mots en rapport avec l'automobile (code 45 imputé).

#### **E- Résultats du codage définitif en NAF rev2**

Au total, 20 938 séquences codées sur les 29 237 du fichier initial (soit 71,6%)

Détail du codage des APE :

position	fréquence	%
0	8 299	28,4
2	12 212	41,8
5	8 726	29,9

## Annexe 8 – Définition des variables annexes utilisées dans Sicore PCS

### DESCRIPTION DU FICHER UTILISE

**STATUT** : statut dans l'emploi (clivage salarié / indépendant).

Les modalités de STATUT pour les règles Sicore sont les suivantes :

- 1 indépendant
- 2 salarié de sa propre entreprise, gérant mandataire, PDG
- 3 salarié
- \* (ou blanc) manquant

**PUB** : statut de l'établissement employeur (distinction public / privé). Cette variable ne concerne que les salariés.

Les modalités de PUB pour les règles Sicore sont les suivantes :

- 1 Etat
- 2 collectivités territoriales, HLM, hôpitaux
- 3 sécurité sociale
- 4 entreprises publiques nationalisées
- 5 privé
- \* (ou blanc) manquant

**SP** : emploi précaire (distinction apprenti / autre). Cette variable ne concerne que les salariés.

Les modalités de SP pour les règles Sicore sont les suivantes :

- 1 apprenti
- \* (ou blanc) manquant

**CPF** : classification professionnelle ou Qualification

Cette variable ne concerne a priori que les salariés (cf. PCS 2003 page 43). Les modalités de CPF pour les règles Sicore sont les suivantes :

!! Modifiées pour la PCS 2003 !!

- 1 manœuvre ou ouvrier spécialisé
- 2 ouvrier qualifié ou ouvrier hautement qualifié ou technicien d'atelier
- 3 agent de maîtrise
- 4 directeur général ou adjoint direct au directeur
- 5 technicien, dessinateur, VRP

- 6 instituteur, assistante sociale, infirmière et autres personnels de catégorie B de la fonction publique
- 7 ingénieur ou cadre
- 8 professeur et personnel de catégorie A de la fonction publique
- 9 employés de bureau, de commerce, agents de service, aides-soignantes, gardiennes d'enfants, personnels de catégorie C ou D de la fonction publique
- 0 autres
- \* (ou blanc) manquant

**FN** : fonction principale

Les modalités de FN pour les règles Sicore sont les suivantes :

!! Modifiées pour la PCS 2003 !!

- 1 production, fabrication, chantiers
- 2 installation, réparation, maintenance
- 3 nettoyage, gardiennage, entretien ménager
- 4 manutention, magasinage, logistique
- 5 secrétariat, saisie, accueil
- 6 gestion, comptabilité
- 7 commerce, vente, technico-commercial
- 8 études, recherche et développement, méthodes
- 0 autres
- \* (ou blanc) manquant

**NBS** : nombre de salariés employés

Cette variable ne concerne a priori que les chefs d'entreprise ou personnes installées à leur compte.

Les modalités de NBS pour les règles Sicore sont les suivantes :

- 1 aucun salarié
- 2 1 ou 2 salarié(s)
- 3 3 à 9 salariés
- 4 10 salariés ou plus
- \* (ou blanc) manquant

**NAF2** et **NAF** : activité principale de l'établissement

Les modalités de NAF2 et NAF pour les règles Sicore sont les suivantes :

- Toutes les modalités possibles de la nomenclature d'activité

- \*\* ou \*\*\*\* (ou blanc) manquant

La variable NAF2 comprend les deux premières positions du code NAF (niveau "division" de la nomenclature). La variable NAF comprend les codes NAF à quatre positions.

Si l'on dispose de la NAF à quatre positions, créer aussi la variable annexe NAF2. Certains codages ne nécessitant que la NAF2.

**S** : sexe

Les modalités de S pour les règles Sicore sont les suivantes :

- 1 masculin
- 2 féminin
- \* (ou blanc) manquant

**T** : taille de l'entreprise

Les modalités de T pour les règles Sicore sont les suivantes :

- 0 de 0 à 9 salariés
- P de 10 à 49 salariés
- M de 50 à 499 salariés
- G 500 salariés et plus
- \* (ou blanc) manquant

**OPA** : orientation des productions agricoles

Les modalités de OPA pour les règles Sicore sont les suivantes :

- 1 polyculture (culture des terres labourables)
- 2 maraîchage ou horticulture
- 3 vigne ou arbres fruitiers
- 4 élevage d'herbivores (bovins, ovins...)
- 5 élevage de granivores (volailles, porcs,..)
- 6 polyculture-élevage
- 7 élevage d'herbivores et de granivores
- 8 autre
- \* (ou blanc) manquant

**DEP** : département

Les modalités de DEP pour les règles Sicore sont les suivantes :

- Tous les numéros de départements métropolitains (pour la Corse, on peut distinguer 2A et 2B ou ne pas distinguer en mettant 20)
- 97 pour les DOM (pas de distinction)
- \*\* (ou blanc) manquant

Le département peut être utile pour la codification de certaines professions d'agriculteurs.



## Annexe 9 – Liste des questions ouvertes

Nom de la variable "ouvert"	Libellé de la variable	Module dans le questionnaire
OMOTET	Pour quelle raison est il/elle parti à l'étranger ? (réponse du tiers)	Questionnaire filtre
OQ6B	Dans quelle spécialité étiez-vous inscrit en 2009-2010 ?	Questionnaire filtre
OQ32	Dans quelle commune habitez-vous lorsque vous êtes entré en 6ème ?	Parcours scolaire
OQ36A	Bac général : Dans quelle série ?	Parcours scolaire
OQ36B	Bac techno/pro : Dans quelle série ?	Parcours scolaire
OQ37	Bac techno/pro : Dans quelle spécialité ?	Parcours scolaire
OQ39	Toujours l'année du bac, avez-vous candidaté à :	Parcours scolaire
OQ39A	Finalement, qu'avez-vous fait immédiatement après la terminale ?	Parcours scolaire
OQ39C1	Quel est ce diplôme ? (pour les sortants du secondaire)	Identification du plus haut diplôme
OQ39C2	Quelle est la spécialité ? (pour les sortants du secondaire)	Identification du plus haut diplôme
OQ39L	Lorsque vous avez arrêté vos études en 2010, possédiez-vous un diplôme ?	Identification du plus haut diplôme
OQ39M	Quelle est la spécialité ?	Identification du plus haut diplôme
OQ39I1	Quelle était ce diplôme ? (pour ceux avec un diplôme de niveau bac+4)	Identification du plus haut diplôme
OQ39I2	Quelle était la spécialité ? (pour ceux avec un diplôme de niveau bac+4)	Identification du plus haut diplôme
OBB39	Bac général : Dans quelle série ?	Identification du plus haut diplôme
OBC39	Bac techno/pro : Dans quelle série ?	Identification du plus haut diplôme
OBD39	Bac techno/pro : Dans quelle spécialité ?	Identification du plus haut diplôme
OETR10	Quelle était cette bourse ou aide financière ?	Séjours à l'étranger
OFP5	Quel diplôme prépariez-vous au moment de ce stage ?	Stages en cours d'études
OQ71A	Année : Quand cet emploi s'est-il terminé ?	Expériences en cours d'études
OTH3	Votre projet professionnel au moment de la soutenance de thèse était de :	Module thèses
OTH4	Avez-vous bénéficié d'une autre rémunération pendant votre thèse ?	Module thèses
OTH6	Quel est le label du laboratoire avec lequel vous avez effectué votre thèse ?	Module thèses
OTH14	Durant la préparation de votre thèse, avez-vous participé à un autre dispositif ?	Module thèses
OST16	Type d'employeur avant entrée en formation	Post-initiaux - Sport
OST17	Quelle était votre profession avant l'entrée en formation ?	Post-initiaux - Sport
OST24A	Bac général : Dans quelle série ? Autre	Post-initiaux - Sport
OST24B	Bac techno/pro : Dans quelle série ? Autre	Post-initiaux - Sport
OST26	Possédez-vous un autre diplôme dans le champs du sport ou de l'animation,	Post-initiaux - Sport
OST261	Possédez-vous un autre diplôme dans le champs du sport ou de l'animation,	Post-initiaux - Sport
OSP16	Type d'employeur avant entrée en formation	Post-initiaux - Santé / Social
OSP24	Série du bac obtenu	Post-initiaux - Santé / Social
OCA6	Dans quel pays ?	Caractéristiques individuelles - Discrimination
OCA26	Discr embauche : Quelle autre raison ?	Caractéristiques individuelles
OCA231	Quelle est cette maladie chronique invalidante ?	Caractéristiques individuelles - AGEFIPH
OCA232	Quelle est cette autre gêne ou limitation ?	Caractéristiques individuelles - AGEFIPH
OCA23F	Quelle est cette autre reconnaissance ou allocation ?	Caractéristiques individuelles - AGEFIPH
OCA23I	Pourquoi n'avez-vous pas obtenu une reconnaissance administrative de votre	Caractéristiques individuelles - AGEFIPH
OEP10	Activité de l'entreprise en clair	Séquence emploi
OEP21	Quel contrat de travail ? (à l'embauche)	Séquence emploi
OEP21B	Quel est le diplôme préparé (à l'embauche) dans le cadre de ce CDD ? Pas de menu	Séquence emploi
OEP19	Profession à l'embauche?	Séquence emploi
OEP27	Quel contrat de travail ? (final)	Séquence emploi
OEP27B	Quel est le diplôme préparé (final) dans le cadre de ce CDD ?	Séquence emploi
OEP40	Profession finale ?	Séquence emploi
OETUP1	Ces études vous préraient à :	Séquence de reprise d'études
OETUA2	Quel est le diplôme que vous préparez actuellement ?	Séquence de reprise d'études
ONEP5D	Pour quelle raison vous n'étiez pas disponible ?	Séquence de non emploi
OMF9A	Quel diplôme avez-vous préparé ? (individu en formation)	Séquence de non emploi
ONEP9AA	Pouvez vous les citer ? (agences d'interim)	Séquence de non emploi

## ANNEXE 10 – Pondérations des extensions

### a) Extension DREES (formations sanitaires et sociales) (PB NUM)



Juillet 2014

#### Complément à la note méthodologique :

#### Pondération spécifique pour l'extension sur les champs des formations sanitaires et sociale

Variable de pondération finale spécifique : POND\_DREES

Le champ de l'extension concerne les sortants de formation sanitaire et sociale sans critères sur l'âge ni sur le fait d'avoir connu une interruption d'étude. Ce champ est plus large que celui du Céreq qui concerne essentiellement les primo sortant, c'est-à-dire les individus sortant pour la première fois du système éducatif, âgés de moins de 35 ans. Les champs étant différents, une pondération spécifique a été calculée pour cette extension.

#### Base de sondage

Les écoles de la santé et du social ne sont pas extraites de la BCE. La collecte a été effectuée auprès de la Drees car les bases réceptionnées nous ont permis d'établir une sélection précise des formations. La nomenclature des formations et leurs codes diplômes nous ont aussi été fournis au même moment. Une relance courrier a été effectuée sur les écoles de la santé et du social, directement par la Drees pour optimiser la collecte des fichiers. Dans la base de sondage sont présents 37 153 individus, parmi eux 21 623 ont été échantillonnés.

**Tableau 1 : Information sur la couverture**

Formation	Code formation	Effectifs collectés dans la base de sondage	Effectifs totaux	Taux de couverture
Aide-soignants	419	10 853	17 974	60 %
Assistant de service social	432	1 446	2 256	64 %
Auxiliaire de puériculture	456	1 484	3 447	43 %
Conseiller en économie sociale et familiale	422	700	1 229	57 %
Educateur de jeunes enfants	435	935	1 214	77 %
Educateur spécialisé	433	2 686	4 244	63 %
Ergothérapeute	423	228	371	61 %

Infirmier	414	14 871	21 641	69 %
Manipulateur d'électro-radiologie	421	222	591	38 %
Masseur-kinésithérapeute	416	1 245	2 015	62 %
Moniteur éducateur	437	1 723	2 487	69 %
Pédicure-Podologue	420	87	449	19 %
Psychomotricien	424	103	515	20 %
Puéricultrice	418	314	1 007	31 %
Sage-femme	415	256	1 256	28 %
Ensemble		37 153	60 696	61 %

**Tableau 2 : Bilan de la collecte sur les établissements de la santé, nombre de répondants, dans le champ et hors champ**

Formation	Effectifs dans le champ	Effectifs hors champ	Ensemble
Aide_soignants	1 361	232	1 593
Assistant de service social	477	101	578
Auxiliaire de puériculture	404	52	456
Conseiller en économie sociale et familiale	182	56	238
Educateur de jeunes enfants	327	34	361
Educateur spécialisé	771	91	862
Ergothérapeute	79	32	111
Infirmier	1 533	394	1 927
Manipulateur d'électro-radiologie	81	7	88
Masseur-kinésithérapeute	287	76	363
Moniteur éducateur	438	74	512
Pédicure-Podologue	20	4	24
Psychomotricien	37	2	39
Puéricultrice	89	17	106
Sage-femme	392	26	418
Ensemble	6 478 <sup>1</sup>	1 198	7 676

### Redressement et calage

Par rapport à la « Génération 2007 », nous avons 1 414 répondant en plus (7 676 contre 6 262).

Par conséquent, il n'apparaît plus nécessaire d'opérer des regroupements de formation tels qu'ils avaient été faits pour la « Génération 2007 ». Deux formations possèdent un nombre limité d'observation, formation de Pédicure-Podologue et de Psychomotricien, pour lesquelles il sera nécessaire d'être prudent dans les interprétations. Les marges de calages utilisées sont donc les effectifs totaux par niveaux de formation présentées dans le tableau 1.

Nous avons réalisé une correction de la non-réponse puis un calage sur marges. Le calage des observations est effectué en partant des poids définis par le produit des poids de couverture, des poids de tirage et des poids de correction de la non-réponse, avec la macro calmar<sup>53</sup>. Après calage, en

<sup>53</sup> <https://www.insee.fr/fr/information/2021902>

enlevant les effectifs hors champ, nous obtenons une population qui représente 50 832 individus dans le champ des formations de la santé sur les 60 696 initialement présents dans la base de sondage.

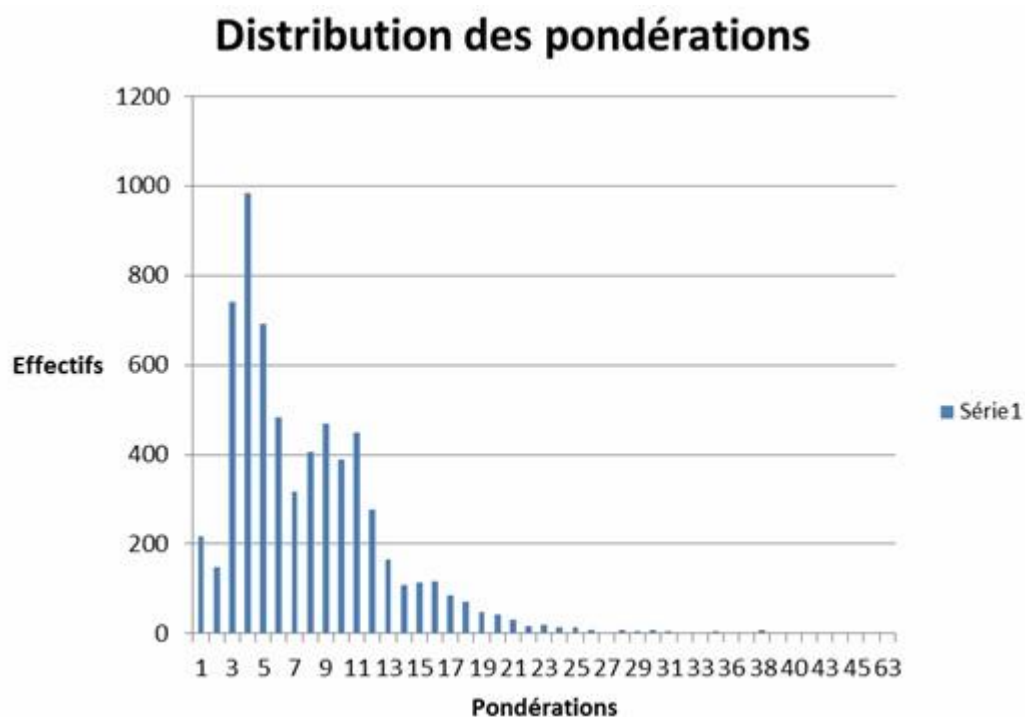
**Tableau 3 : structure des répondants avec et sans pondération**

	Effectifs non pondérés	%	Effectifs pondérés	%
Champ santé	6 478	100	50 832	100
dont champ Céreq	3 797	59	28 937	57

**Tableau 4 : Structure des répondants selon les niveaux de formation avec et sans pondération**

Formation	Code formation	Effectifs non pondérés	%	Effectifs pondérés	%
Infirmier	414	1 533	24	17 281	34
Sage-femme	415	392	6	1 089	2
Masseur-kinésithérapeute	416	287	4	1 637	3
Puéricultrice	418	89	1	849	2
Aide-soignants	419	1 361	21	15 348	30
Pédicure-Podologue	420	20	0	379	1
Manipulateur d'électro-radiologie	421	81	1	512	1
Conseiller en économie sociale et familiale	422	182	3	930	2
Ergothérapeute	423	79	1	284	1
Psychomotricien	424	37	1	495	1
Assistant de service social	432	477	7	1 872	4
Educateur spécialisé	433	771	12	3 819	8
Educateur de jeunes enfants	435	327	5	1 104	2
Moniteur éducateur	437	438	7	2 163	4
Auxiliaire de puériculture	456	404	6	3 070	6
Ensemble		6 478	100	50 832	100

## Structure des poids finaux



<sup>i</sup> Le nombre de sortants dans le champ diffère légèrement par rapport aux notes précédemment envoyées : les sortants ayant un diplôme d'état de conseiller en économie sociale et familiale, qui correspondent à 19 répondants dans le champ n'avaient pas été pris en compte.

## b) Extension SPORT (formations sport et animations)



Juillet 2014

### Complément à la note méthodologique :

#### Pondération spécifique pour l'extension sur les champs des formations sport et animation

Variable de pondération finale spécifique : POND_SPORT
--

Le champ de l'extension sur les formations des secteurs du sport et de l'animation est constitué par les individus diplômés ayant suivi en 2009-2010 une formation dans un de ces secteurs et ayant terminé leurs études en 2009-2010. Le champ de l'extension est donc plus large que le champ du Céreq puisqu'il inclut les individus post-initiaux, sans limite d'âge, qui avaient interrompu une première fois leurs études avant 2009-2010.

Parmi les 23 DRJS sollicitées par le Céreq, 17 ont renvoyé les informations demandées concernant les sortants de leur établissement. La base de sondage contient au final 11 680 individus diplômés de DRJS. En parallèle, le ministère des Sports a transmis au Céreq un fichier national des effectifs de diplômés en 2010.

En comparant les effectifs issus des fichiers bruts collectés par le Céreq et ceux du fichier national transmis par le ministère des Sports, on estime le taux de couverture de notre base de sondage par rapport au champ des formations du sport et de l'animation, à environ 65 %. Le nombre de diplômés est de 17 953 selon le ministère. Tous les diplômés présents dans la base de sondage appartenant à l'extension du sport et de l'animation, soit 12 477 individus, ont été échantillonnés.

Parmi les sortants appartenant au champ de l'extension et présents dans la base de sondage, aucun n'a suivi de formation BEATEP (Brevet d'état d'animateur technicien de l'éducation populaire et de la jeunesse) ni de formation DEDPAD (Diplôme d'Etat Directeur de Projet d'Animation et de développement).

Les regroupements des formations sont présentés dans le tableau suivant :

**Codification des formations**

Libellé en clair	Code du libellé
Brevet d'aptitude professionnelle d'assistant animateur technicien de la jeunesse et du sport	BAP
Brevet d'état d'éducateur sportif 3 <sup>e</sup> degré	BEE
Brevet d'état d'éducateur sportif 2 <sup>d</sup> degré	BEE
Brevet d'état d'éducateur sportif 1 <sup>er</sup> degré	BEE
BEES sans précision du degré, ressemble à du premier degré	BEE
Perfectionnement sportif, sans doute du BEES second degré	BEE
Brevet professionnel de la jeunesse, de l'éducation populaire et du sport, spécialité animation	BPJ
Brevet professionnel de la jeunesse, de l'éducation populaire et du sport, spécialité sport	BPJ
Spécialité animation, sans précision du diplôme. Sans doute du brevet professionnel	BPJ
Spécialité sport, sans précision du diplôme. Sans doute du brevet professionnel	BPJ
Diplôme d'Etat relatif aux fonctions d'animation	DEF
Diplôme d'Etat supérieur de la jeunesse, de l'éducation populaire et du sport	DEJ
Diplôme d'Etat de la jeunesse, de l'éducation populaire et du sport	DEJ

**Taux de couvertures de la base de sondage**

Formation	Nombre de diplômés collectés	Effectif total	Taux de couverture
BAP	513	836	61 %
BEE	3 426	4 693	73 %
BPJ	5 470	8 690	63 %
DEF	28	86	33 %
DEJ	3 040	3 648	83 %
Ensemble	12 477	17 953	69 %

Le nombre total de diplômés établi à partir des fichiers transmis par les DRJS, à savoir 12 477 diplômés, dépasse ici les 11 680 individus présents dans la base de sondage, du fait que des sortants des formations de la jeunesse et des sports peuvent avoir plusieurs diplômes. Par la suite, nous ne considérons qu'un diplôme par sortants.

Résultat de la collecte**Résultats de la collecte par niveau de formation**

Formation	Effectifs répondant dans le champ	Effectifs répondant hors champ	Ensemble
BAP	68	30	98
BEE	415	311	726
BPJ	847	277	1 124
DEF	6	4	10
DEJ	371	254	625
Ensemble	1 707	876	2 583

Le niveau de formation DEF (Diplôme d'état relatif aux fonctions d'animation) contient très peu d'observations, nous procédons donc à un regroupement des diplômes DEF et DEJ dans une modalité plus générale pour les diplômes d'état (DE\_) en vue des étapes de calage. Les marges de calages sont donc les suivantes :

### Marges de calage

Marges	Nombre de répondants
<b>BAP</b> Brevet d'aptitude professionnelle d'assistant animateur technicien de la jeunesse et du sport	836
<b>BEE</b> Brevet d'état d'éducateur	4 693
<b>BPJ</b> Brevet professionnel de la jeunesse	8 690
<b>DEF +DEJ</b> Diplôme d'état relatif aux fonctions d'animation + Diplôme d'Etat de la jeunesse, de l'éducation populaire et du sport	3 734
Ensemble	17 953

### Correction de la non réponse et calage

Nous avons réalisé une correction de la non-réponse puis un calage sur marges. Le calage des observations est effectué en partant des poids définis par le produit des poids de couverture, des poids de tirage et des poids de correction de la non-réponse, avec la macro calmar<sup>54</sup>. Après calage, les répondants dans le champ représentent 11 585 sortants sur les 17 953 individus initialement présents dans la base de sondage.

### Effectifs pondérés et non pondérés des répondants dans le champ jeunesse et sport

Classement des questionnaires	Effectifs non pondérés	Effectif pondérés
champ de la jeunesse et des sports	1707	11 585
<i>dont champ Céreq</i>	535	3 992

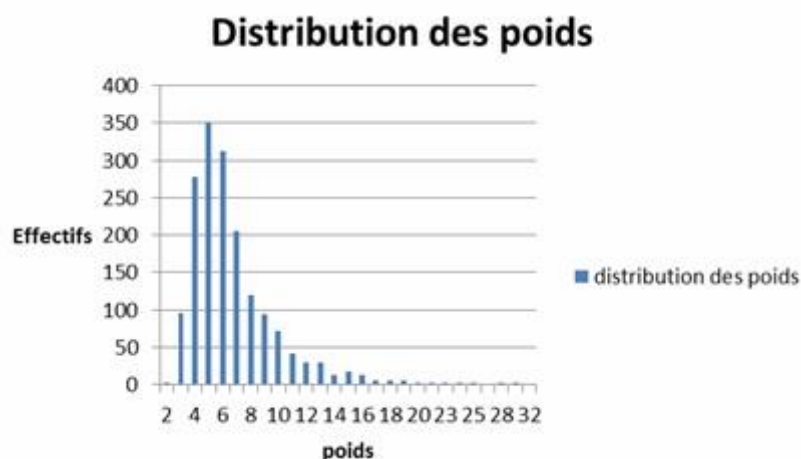
### Structure des sortants par niveau de formation

Marges	Effectifs non pondérés	%	Effectifs	%
Brevet d'aptitude professionnelle d'assistant animateur technicien de la jeunesse et du sport	68	4	524	5
Brevet d'état d'éducateur	415	24	2 551	22
Brevet professionnel de la jeunesse	847	50	6 334	55
Diplôme d'état relatif aux fonctions d'animation +Diplôme d'Etat de la jeunesse, de l'éducation populaire et du sport	377	22	2 176	19
Ensemble	1 707	100	11 585	100

<sup>54</sup> <https://www.insee.fr/fr/information/2021902>



## Structure des poids après pondération



### c) Extension DGAFP (formations de la fonction publique)



Juillet 2014

#### Complément à la note méthodologique :

#### Pondération spécifique pour l'extension sur les champs des formations de la fonction publique

Variable de pondération finale spécifique : POND\_DGAFP

Le champ de l'extension DGAFP concerne les sortants des écoles de la fonction publique sans critères sur l'âge ni sur le fait d'avoir connu une interruption d'étude. Ce champ est plus large que celui du Céreq qui concerne essentiellement les primo sortant, c'est-à-dire les individus sortant pour la première fois du système éducatif, âgés de moins de 35 ans.

Les établissements de la fonction publique ne sont renseignés que pour moitié dans la BCE. Une liste a donc été demandée à la DGAFP en vue de les intégrer dans le système de collecte. Cette collecte a été intégrée à la collecte générale, avec une insistance plus particulière sur ces établissements.

#### Résultats de l'enquête

Sur les 64 établissements de la liste envoyée par la DGAFP, 47 ont transmis les informations concernant leur établissement. Malgré l'extension du champ de l'enquête génération pour cette génération 2010, il n'y a pas de sortants des Doms dans les bases de sondages constituées.

**Etablissements qui ont participé à la collecte :**

MATRICULE	NOM	LOCALISATION	Type d'école
013F01	ENTE	AIX-EN-PROVENCE CEDEX 3	ECOLE DE LA DGAFP - SUPPLEMENT
013F02	ECOLE NAT DE POLICE	FOS-SUR-MER CEDEX	ECOLE DE LA DGAFP - SUPPLEMENT
0211660R	ECOLE NATIONALE DES GREFFES	DIJON CEDEX	ECOLE DE LA DGAFP - SUPPLEMENT
024F01	ECOLE NAT DE POLICE	PÉRIGUEUX	ECOLE DE LA DGAFP - SUPPLEMENT
0251889T	ECOLE NAT DE POLICE	MONT BÉLIARD	ECOLE DE LA DGAFP - SUPPLEMENT
0290125D	ENSI ETUD TECHN ARMEMENT BREST	BREST	DGA
0300063F	ENS TECH INDUS ET MINES ALES	ALES CEDEX	ECOLE DE LA DGAFP - SUPPLEMENT
0310153Y	ECOLE NAT VETERINAIRE TOULOUSE	TOULOUSE CEDEX 3	ECOLE DE FORMATION AGRICOLE
0311256X	EC NAT AVIATION CIV TOULOUSE	TOULOUSE CEDEX 04	ECOLE DE LA DGAFP - SUPPLEMENT
031F04	CENTRE FORM POLICE NAT	TOULOUSE CEDEX	ECOLE DE LA DGAFP - SUPPLEMENT
0332239E	ECOLE NATIONALE MAGISTRATURE	BORDEAUX CEDEX	ECOLE DE LA DGAFP - SUPPLEMENT
0342241B	EC NAT CONC CONS REP FRAUDES	MONTPELLIER CEDEX 9	ECOLE DE LA DGAFP - SUPPLEMENT
0352554L	ECOLE NAT DE POLICE	SAINT MALO CEDEX	ECOLE DE LA DGAFP - SUPPLEMENT
035F01	ENSAI	BRUZ	ECOLE DE LA DGAFP - SUPPLEMENT
0441618V	INST REGIONAL D'ADMINISTRATION	NANTES CEDEX 3	ECOLE ADMINISTRATION PUBLIQUE
0442205H	ENS TECH INDUST MINES NANTES	NANTES CEDEX 3	ECOLE DE LA DGAFP - SUPPLEMENT
0442674T	ECO. NAT. VET. AGRO ALIM	NANTES CEDEX 3	ECOLE DE FORMATION AGRICOLE
0492189N	CFR ANGERS- INHP	ANGERS CEDEX 01	ECOLES INGENIEURS
0540128W	EC NAT SUP MINES DE NANCY	NANCY CEDEX	ECOLES INGENIEURS
0572647T	INST REGIONAL D'ADMINISTRATION	METZ	ECOLES ADMINISTRATION PUBLIQUE
0593648V	INST REGIONAL D'ADMINISTRATION	LILLE CEDEX	ECOLES ADMINISTRATION PUBLIQUE
0596695G	ECOLE NATIONALE DES DOUANES	TOURCOING CEDEX	ECOLE DE LA DGAFP - SUPPLEMENT
0596802Y	ECOLE NAT DE POLICE	ROUBAIX CEDEX 1	ECOLE DE LA DGAFP - SUPPLEMENT
059F01	ENTE	VALENCIENNES CEDEX	ECOLE DE LA DGAFP - SUPPLEMENT
0631786Z	EC NAT ING TRAV AGRI CLERMT -FD	LEMPDES	ECOLES INGENIEURS
0670189S	EN GENIE EAU ENVIRON STRASBG	STRASBOURG CEDEX	ECOLES INGENIEURS
0672678X	EC NAT.D'ADMINISTRATION	STRASBOURG CEDEX	ECOLES ADMINISTRATION PUBLIQUE
0690189F	EC NATIONALE SUP DE LA POLICE	ST CYR AU MONT D OR	ECOLE DE LA DGAFP - SUPPLEMENT
0690193K	INSTITUT ENS SUP RECH AGRONOM	MARCY LETOILE	ECOLE DE FORMATION AGRICOLE
0691861Y	INST REGIONAL D'ADMINISTRATION	VILLEURBANNE CEDEX	ECOLES ADMINISTRATION PUBLIQUE
0693183K	INFOMA	CORBAS	ECOLE DE FORMATION AGRICOLE
0694123G	ECOLE NORMALE SUP LYON	LYON CEDEX 07	ECOLES NORMALES SUPERIEURES
069F01	ENSSIB	VILLEURBANNE CEDEX	ECOLE DE LA DGAFP - SUPPLEMENT
069F02	CENTRE FORM POLICE NAT	CHASSIEU	ECOLE DE LA DGAFP - SUPPLEMENT
069F11	INTEFP	MARCY L'ÉTOILE	ECOLE DE LA DGAFP - SUPPLEMENT
0753455Y	ECOLE NORMALE SUPERIEURE PARIS	PARIS CEDEX 05	ECOLES NORMALES SUPERIEURES
0753503A	EN GEN RURAL EAUX FORET PARIS	PARIS CEDEX 15	ECOLES INGENIEURS
0755222U	CENTRE DES HAUTES ETUDES	PARIS	ECOLES ARCHITECTURES ET ARTISTIQUES
075F09	CEFIL	LIBOURNE	ECOLE DE LA DGAFP - SUPPLEMENT
0763238G	ECOLE NAT DE POLICE	OISSEL	ECOLE DE LA DGAFP - SUPPLEMENT
0763402K	ECOLE NATIONALE DES DOUANES	ROUEN CEDEX	ECOLE DE LA DGAFP - SUPPLEMENT
0772316Z	EC.SUP.INSPECTEURS POLICE NAT.	CANNES ECLUSE	ECOLE DE LA DGAFP - SUPPLEMENT
0772517T	EN PONTS CHAUSS CHAMPS/MARNE	CHAMPS SUR MARNE	ECOLES INGENIEURS
0811200P	ENS TECH INDUS MINES ALBI	ALBI CT CEDEX 09	ECOLE DE LA DGAFP - SUPPLEMENT
092F01	ENSAE	MALAKOFF CEDEX	ECOLE DE LA DGAFP - SUPPLEMENT
0940608A	ECOLE NAT VETERINAIRE ALFORT	MAISONS ALFORT CEDEX	ECOLE DE FORMATION AGRICOLE
7200090K	INST REGIONAL D ADMINISTRATION	BASTIA	ECOLES ADMINISTRATION PUBLIQUE

**Etablissements qui n'ont pas participé à la collecte :**

MATRICULE	NOM	LOCALISATION	Type d'école
017F01	ECOLE NATIONALE DES DOUANES	LA ROCHELLE CEDEX 1	ECOLE DE LA DGAFP - SUPPLEMENT
030F01	ECOLE NAT DE POLICE	NIMES CEDEX 1	ECOLE DE LA DGAFP - SUPPLEMENT
0312069F	EC NAT METEOROLOGIE TOULOUSE	TOULOUSE CEDEX 1	ECOLE DE LA DGAFP - SUPPLEMENT
0350095N	EC HTES ETUD SANTE PUB RENNES	RENNES CEDEX	ECOLES INGENIEURS
0420094Z	ENS MINES ST ETIENNE	ST ETIENNE CEDEX 2	ECOLE DE LA DGAFP - SUPPLEMENT
047F01	ENAP	AGEN CEDEX 9	ECOLE DE LA DGAFP - SUPPLEMENT
0511986G	ECOLE NAT DE POLICE	REIMS CEDEX	ECOLE DE LA DGAFP - SUPPLEMENT
0590342B	ENS TECH INDUST MINES DOUAI	DOUAI CEDEX	ECOLE DE LA DGAFP - SUPPLEMENT
059F10	ÉCOLE NAT DE LA PROT. JUD. DE LA JEUN.	ROUBAIX CEDEX 1	ECOLE DE LA DGAFP - SUPPLEMENT
063F01	ECOLE NAT DE POLICE	CHAT EL GUYON	ECOLE DE LA DGAFP - SUPPLEMENT
0692566P	EN TRAV PUBL ETAT VX EN VELIN	VAULX EN VELIN CEDEX	ECOLE DE LA DGAFP - SUPPLEMENT
0753493P	ENS MINES PARIS	PARIS CEDEX 06	ECOLE DE LA DGAFP - SUPPLEMENT
0753510H	ENS TELECOMMUNICATIONS PARIS	PARIS CEDEX 13	ECOLES INGENIEURS
075F01	INSTITUT NATIONAL DU PAT. (INP)	PARIS	ECOLE DE LA DGAFP - SUPPLEMENT
0772496V	EN SC GEOGRAPH CHAMPS/MARNE	CHAMPS SUR MARNE CEDEX	ECOLE DE LA DGAFP - SUPPLEMENT
093F01	ENFIP	NOISY-LE-GRAND CEDEX	ECOLE DE LA DGAFP - SUPPLEMENT
0940607Z	ECOLE NORMALE SUP CACHAN	CACHAN	ECOLES NORMALES SUPERIEURES

Tous les individus du champ DGAFP ont été tirés dans l'échantillon, soit 7 045 individus. Sur ces 7 045 individus échantillonnés, 2 285 ont répondu indépendamment de leur appartenance au champ de la DGAFP. Ces répondants, dans le champ DGAFP et hors champ, se répartissent de la manière suivante selon les établissements de formation :

## Structure des répondants par établissement

MATRICULE	NOM	LOCALISATION	Effectifs dans échantillon	Effectifs des répondants dans le champ	Effectifs des répondants hors champ
013F01	ENTE	AIX-EN-PROVENCE CEDEX 3	178	69	1
013F02	ECOLE NAT DE POLICE	FOS-SUR-MER CEDEX	276	43	3
0211660R	ECOLE NATIONALE DES GREFFES	DIJON CEDEX	471	148	25
024F01	ECOLE NAT DE POLICE	PÉRIGUEUX	167	40	2
0251889T	ECOLE NAT DE POLICE	MONTBÉLIARD	118	29	1
0290125D	ENSI ETUD TECHN ARMEMENT BREST	BREST	167	27	13
0300063F	ENS TECH INDUS ET MINES ALES	ALES CEDEX	172	25	12
0310153Y	ECOLE NAT VETERINAIRE TOULOUSE	TOULOUSE CEDEX 3	14	3	1
0311256X	EC NAT AVIATION CIV TOULOUSE	TOULOUSE CEDEX 04	216	50	11
031F04	CENTRE FORM POLICE NAT	TOULOUSE CEDEX	73	20	1
0332239E	ECOLE NATIONALE MAGISTRATURE	BORDEAUX CEDEX	201	0	1
0342241B	EC NAT CONC CONS REP FRAUDES	MONTPELLIER CEDEX 9	100	17	0
0352554L	ECOLE NAT DE POLICE	SAINT MALO CEDEX	154	51	2
035F01	ENSAI	BRUZ	119	48	12
0441618V	INST REGIONAL D'ADMINISTRATION	NANTES CEDEX 3	130	53	5
0442205H	ENS TECH INDUST MINES NANTES	NANTES CEDEX 3	128	37	10
0442674T	ECO, NAT, VET, AGRO ALIM	NANTES CEDEX 3	16	2	7
0492189N	CFR ANGERS- INHP	ANGERS CEDEX 01	114	50	10
0540128W	EC NAT SUP MINES DE NANCY	NANCY CEDEX	150	45	11
0572647T	INST REGIONAL D'ADMINISTRATION	METZ	132	65	2
0593648V	INST REGIONAL D'ADMINISTRATION	LILLE CEDEX	130	46	4
0596695G	ECOLE NATIONALE DES DOUANES	TOURCOING CEDEX	134	9	0
0596802Y	ECOLE NAT DE POLICE	ROUBAIX CEDEX 1	191	36	4
059F01	ENTE	VALENCIENNES CEDEX	78	27	4
0631786Z	EC NAT ING TRAV AGRI CLERMT-FD	LEMPDES	110	45	4
0670189S	EN GENIE EAU ENVIRON STRASBG	STRASBOURG CEDEX	131	60	8
0672678X	EC NAT,D'ADMINISTRATION	STRASBOURG CEDEX	81	33	0
0690189F	EC NATIONALE SUP DE LA POLICE	ST CYR AU MONT D OR	58	17	0
0690193K	INSTITUT ENS SUP RECH AGRONOM	MARCY L ETOILE	68	19	14
0691861Y	INST REGIONAL D'ADMINISTRATION	VILLEURBANNE CEDEX	70	27	2
0693183K	INFOMA	CORBAS	84	25	3
0694123G	ECOLE NORMALE SUP LYON	LYON CEDEX 07	667	48	287
069F01	ENSSIB	VILLEURBANNE CEDEX	106	42	6
069F02	CENTRE FORM POLICE NAT	CHASSIEU	98	21	2
069F11	INTEFP	MARCY L'ÉTOILE	102	6	0
0753455Y	ECOLE NORMALE SUPERIEURE PARIS	PARIS CEDEX 05	208	14	79
0753503A	EN GEN RURAL EAUX FORET PARIS	PARIS CEDEX 15	102	25	4
0755222U	CENTRE DES HAUTES ETUDES	PARIS	54	13	7
075F09	CEFIL	LIBOURNE	96	2	0
0763238G	ECOLE NAT DE POLICE	OISSEL	493	126	14
0763402K	ECOLE NATIONALE DES DOUANES	ROUEN CEDEX	98	2	3
0772316Z	EC,SUP,INSPECTEURS POLICE NAT,	CANNES ECLUSE	99	17	1
0772517T	EN PONTS CHAUSS CHAMPS/MARNE	CHAMPS SUR MARNE	272	26	32
0811200P	ENS TECH INDUS MINES ALBI	ALBI CT CEDEX 09	141	35	15
092F01	ENSAE	MALAKOFF CEDEX	128	39	13
0940608A	ECOLE NAT VETERINAIRE ALFORT	MAISONS ALFORT CEDEX	22	2	5
7200090K	INST REGIONAL D ADMINISTRATION	BASTIA	128	59	1
ENSEMBLE			7045	1643	642

Correction de la non réponse et calage

Nous avons réalisé une correction de la non-réponse puis un calage sur marges. Le calage des observations est effectué en partant des poids définis par le produit des poids de couverture, des poids de tirage et des poids de correction de la non-réponse, avec la macro calmar.<sup>55</sup>

Cependant, nous ne disposons pas des effectifs des écoles non répondantes, nous effectuons donc le calage sur le champ des écoles de la fonction publique ayant participé à la collecte. Nous construisons des marges de calage à partir d'une classification mobilisant les variables niveau de formation des écoles, taux de réponse par écoles, taux de hors champ, catégories sociales des sortants, salaire des sortants. Nous choisissons une partition en cinq classes à partir du critère du saut d'inertie.

La classification des établissements est présentée dans le tableau suivant :

**Classification des établissements :**

Marges	Code établissement	Nom
1	0290125D	ENSI_ETUD_TECHN_ARMEMENT_BREST
1	0300063F	ENS_TECH_INDUS_ET_MINES_ALES
1	0310153Y	ECOLE_NAT_VETERINAIRE_TOULOUSE
1	0311256X	EC_NAT_AVIATION_CIV_TOULOUSE ENSAI
1	035F01	INST_REGIONAL_D'ADMINISTRATION
1	0441618V	ENS_TECH_INDUST_MINES_NANTES
1	0442205H	CFR_ANGERS-_INHP
1	0492189N	EC_NAT_SUP_MINES_DE_NANCY
1	0540128W	INST_REGIONAL_D'ADMINISTRATION
1	0572647T	INST_REGIONAL_D'ADMINISTRATION
1	0593648V	EC_NAT_ING_TRAV_AGRI_CLERMT-FD
1	0631786Z	EN_GENIE_EAU_ENVIRON_STRASBG
1	0670189S	INSTITUT_ENS_SUP_RECH_AGRONOM
1	0690193K	INST_REGIONAL_D'ADMINISTRATION ENSSIB
1	0691861Y	EN_GEN_RURAL_EAUX_FORET_PARIS
1	069F01	CENTRE_DES_HAUTES_ETUDES
1	0753503A	EN_PONTS_CHAUSS_CHAMPS/MARNE
1	0755222U	ENS_TECH_INDUS_MINES_ALBI
1	0772517T	ENSAE
1	0811200P	INST_REGIONAL_D_ADMINISTRATION
1	092F01	ECOLE_NAT_DE_POLICE
1	7200090K	ECOLE_NAT_DE_POLICE
2	013F02	ECOLE_NAT_DE_POLICE
2	024F01	CENTRE_FORM_POLICE_NAT
2	0251889T	ECOLE_NAT_DE_POLICE
2	031F04	ECOLE_NAT_DE_POLICE
2	0352554L	CENTRE_FORM_POLICE_NAT
2	0596802Y	ECOLE_NAT_DE_POLICE
2	069F02	
2	0763238G	

<sup>55</sup> <https://www.insee.fr/fr/information/2021902>

2	0763402K	ECOLE_NATIONALE_DES_DOUANES
3	0342241B	EC_NAT_CONC_CONS_REP_FRAUDES
3	0596695G	ECOLE_NATIONALE_DES_DOUANES
3	0672678X	EC_NAT_D'ADMINISTRATION
3	0690189F	EC_NATIONALE_SUP_DE_LA_POLICE
3	069F11	INTEFP
4	013F01	ENTE
4	0211660R	ECOLE_NATIONALE_DES_GREFFES
4	059F01	ENTE
4	0693183K	INFOMA
4	075F09	CEFIL
4	0772316Z	EC_SUP_INSPECTEURS_POLICE_NAT_
5	0332239E	ECOLE_NATIONALE_MAGISTRATURE
5	0442674T	ECO__NAT__VET__AGRO__ALIM
5	0694123G	ECOLE_NORMALE_SUP_LYON
5	0753455Y	ECOLE_NORMALE_SUPERIEURE_PARIS
5	0940608A	ECOLE_NAT_VETERINAIRE_ALFORT

Après calage, nous obtenons une population qui représente 5 235 individus dans le champ DGAFP sur les 7 045 initialement présents dans la base de sondage :

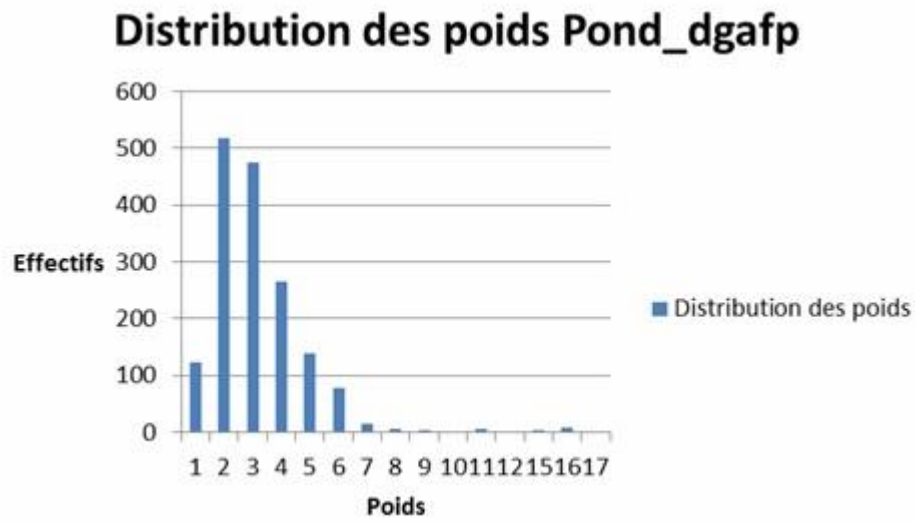
#### Structures des répondants dans le champ selon les marges de calages

Marges	Effectifs non pondérés	%	Effectifs pondérés	%
1	839	51	2 144	41
2	368	22	1 543	29
3	82	5	475	9
4	288	18	928	18
5	66	4	173	3
Ensemble	1 643	100	5 264	100

#### Structure des répondants avec et sans pondération

	Effectifs non pondérés	%	Effectifs pondérés	%
champ dgafp	1643	100	5 264	100
dont champ Céreq	445	27	1 337	25

### Distribution des poids après calage pour les observations dans le champ

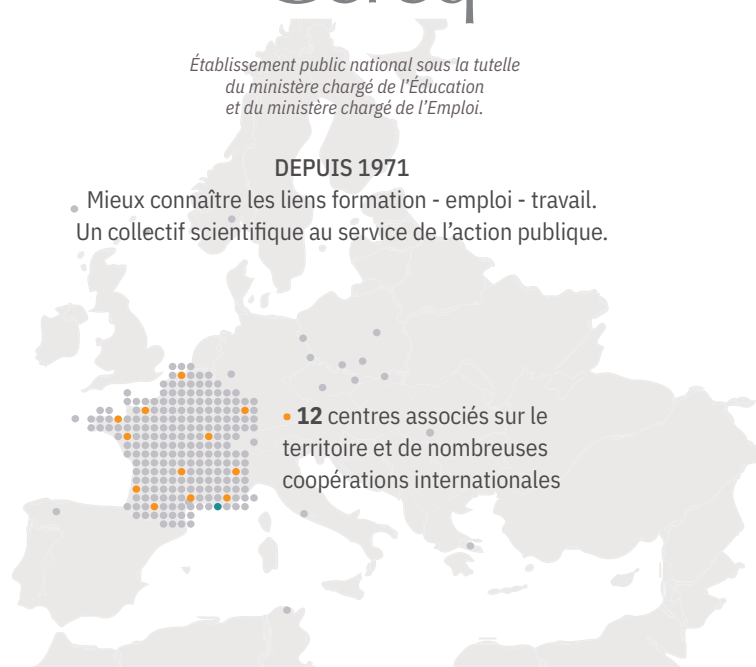


# Céreq

*Établissement public national sous la tutelle  
du ministère chargé de l'Éducation  
et du ministère chargé de l'Emploi.*

## DEPUIS 1971

Mieux connaître les liens formation - emploi - travail.  
Un collectif scientifique au service de l'action publique.




• 12 centres associés sur le territoire et de nombreuses coopérations internationales

 **+ d'infos**  
et tous les travaux

À explorer  
[www.cereq.fr](http://www.cereq.fr)



 **+ de 600 publications**  
Accessibles librement