



Appariement expérimental de l'enquête Génération 2017 avec les fichiers annuels des apprenants de 2005 à 2016 de la DEPP : des résultats prometteurs

Stéphane Jugnot

► To cite this version:

Stéphane Jugnot. Appariement expérimental de l'enquête Génération 2017 avec les fichiers annuels des apprenants de 2005 à 2016 de la DEPP : des résultats prometteurs. 2023. halshs-04077695

HAL Id: halshs-04077695

<https://shs.hal.science/halshs-04077695>

Preprint submitted on 21 Apr 2023

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.



Distributed under a Creative Commons Attribution - NoDerivatives 4.0 International License



RÉPUBLIQUE
FRANÇAISE

*Liberté
Égalité
Fraternité*

Céreq WORKING PAPER

19
2023

Appariement
expérimental de
l'enquête Génération
2017 avec les
fichiers annuels des
apprenants de 2005
à 2016 de la DEPP :
des résultats
prometteurs

DOCUMENT DE TRAVAIL

ENQUÊTE GÉNÉRATION 2017
MÉTHODOLOGIE D'ENQUÊTE
FRANCE
COMPARAISON
ÉDUCATION
FORMATION
INSERTION

Stéphane JUGNOT
Céreq > Direction scientifique

Sommaire

Synthèse	2
1. Les données mobilisées et les objectifs du rapprochement expérimental.....	6
1.1 Une présentation succincte des données mobilisées.....	6
1.2 Les objectifs du rapprochement expérimental	11
2. L'organisation pratique de l'appariement	12
2.1 Le rôle clef de l'INE	12
2.2 Un individu sur cinq de l'échantillon n'a pas d'INE valide.....	15
2.3 Traitement des rares cas d'individus de l'échantillon présents plusieurs fois la même année dans les fichiers des apprenants	19
3. Quels répondants sont retrouvés dans les fichiers des apprenants ?.....	21
4. Reconstruire les étapes principales des parcours d'inscription avec les données administratives	31
4.1 Construire une variable synthétique sur la classe d'inscription	31
4.2 Repérer les classes d'inscription successives des répondants	34
4.3 Repérer les redoublements dans les données administratives et comparer avec ce que dit l'enquête.....	41
4.4 Repérer les inscriptions en 3 ^{ème} et l'orientation en fin de 3 ^{ème} dans les données administratives et comparer avec ce que dit l'enquête	43
4.5 Repérer les inscriptions et les réussites à certains diplômes dans les données administratives et comparer avec ce que dit l'enquête	46
4.6 Examiner de façon plus précise le cas du baccalauréat.....	49
5. Quelques informations complémentaires issues des données administratives.....	52
5.1 La nationalité de l'élève est renseignée de façon hétérogène selon la source	52
5.2 La catégorie sociale des parents n'est disponible que partiellement	53
5.3 Les département de résidence en 6 ^{ème} et au baccalauréat	54
5.4 Autres informations complémentaires illustratives.....	55
6. La non-réponse	57
ANNEXE 1.A- Nomenclature transversale d'identification des classes	58
ANNEXE 1.B – Agrégation <i>ad hoc</i> pour définir les classes successives d'inscription (Programme SAS)	72
ANNEXE 1.C – Répartition des effectifs des fichiers des apprenants transmis par classes, selon notre regroupement <i>ad hoc</i>	74
ANNEXE 2 – Tableaux selon les niveaux de sortie (NIVSOR).....	77
ANNEXE 3 – Programme SAS	95
Bibliographie	143

Synthèse

Depuis le début des années 1990, les enquêtes « Génération », labellisées « Statistique publique » au sens de la loi du 7 juin 1951, permettent au Centre d'études et de recherches sur les qualifications (Céreq) de documenter les premières années de vie active des jeunes après leur sortie de formation initiale. Ces enquêtes, contraintes dans leur durée et par leur budget, abordent le parcours scolaire des jeunes et décrivent leur parcours vers l'emploi. La dernière enquête réalisée porte sur la « Génération 2017 », collectée auprès des jeunes sortis de formation initiale au cours ou à l'issue de l'année scolaire 2016/2017, qui ont été interrogés plus de trois ans après leur sortie, entre fin 2020 et début 2021¹.

Depuis quelques années, la facilitation et le développement des utilisations des données administratives à des fins de recherche et de production statistique créent de nouvelles opportunités. Parmi celles-ci, la mise en œuvre d'un identifiant national des élèves (INE), unique et certifié, ainsi que les investissements réalisés par la DEPP² sur ses données administratives. Ensemble, ils devraient conduire progressivement à mieux suivre les parcours scolaires à partir des données administratives issues des processus de gestion. Comme certaines informations indispensables à la bonne compréhension du lien formation-emploi sont absentes des données administratives, il est vain d'imaginer que les données administratives puissent se substituer totalement aux données d'enquête, mais un rapprochement de ces données avec l'enquête Génération pourrait permettre de réduire les coûts et la charge d'enquête qui pèse sur les enquêtés en raccourcissant la durée d'interrogation. Certaines informations permettraient aussi d'enrichir les informations collectées.

Dans cette perspective, un appariement expérimental a été réalisé entre l'enquête Génération 2017 et les fichiers annuels des apprenants construits par la DEPP. Il vise à examiner les potentialités de ces nouvelles opportunités et à mieux connaître une partie des données produites par la DEPP.

De façon plus précise, la DEPP construit, pour chaque année scolaire, un fichier annuel des apprenants à partir de différentes sources administratives :

- des *extractions du système d'information « Scolarité »*, désigné ensuite sous le terme « SYSCA »³. Elles permettent de disposer d'informations sur les élèves inscrits au cours de l'année scolaire dans l'ensemble des établissements du second degré publics ou privés sous contrat du ministère de l'Éducation nationale (les élèves scolarisés dans le second degré dans les établissements relevant du ministère de l'Agriculture, dans les lycées militaires et dans la majorité des établissements privés hors contrat ne sont pas couverts) ;
- des *extractions du système d'information de la formation des apprentis* (SIFA). Depuis 2008, elles permettent de fournir des informations sur les apprentis inscrits au 1^{er} janvier dans ces formations ;
- des *extractions des systèmes d'information développés pour gérer les inscriptions et les résultats aux examens*. Ils permettent de proposer de compléter les informations sur les inscrits par leurs résultats au diplôme national du brevet, au baccalauréat, aux CAP et aux BTS, sans couvrir les spécialités agricoles.

La construction de ces fichiers est complexe et nécessite de nombreux traitements, par exemple, pour repérer les éventuels doublons ou associer les informations relatives aux examens aux bons élèves. Les informations disponibles dans SYSCA et SIFA ne sont pas non plus strictement les mêmes et sont évolutives.

Pour le rapprochement expérimental, des extraits des fichiers annuels des apprenants 2005 à 2016 ont été mobilisés pour essayer de retrouver des informations sur les parcours scolaires depuis le collège pour la plupart des répondants à l'enquête Génération 2017, quel que soit leur niveau de sortie du système éducatif. Ce document de travail présente les résultats. En résumé, ils sont prometteurs, mais ils ne permettent pas,

¹ La phase de collecte a été décalée de quelques mois par rapport au calendrier habituel en raison de la survenue de la crise sanitaire de la Covid 19 et des restrictions sur les libertés publiques mises en place pour la contenir à partir de mars 2020.

² La Direction de l'Évaluation, de la Prospective et de la Performance (DEPP) est le service statistique du ministère chargé de l'Éducation nationale. Elle est chargée de la conception, de la gestion et de l'exploitation du système d'information statistique sur l'enseignement. Elle a aussi la responsabilité des répertoires et nomenclatures utilisés dans les systèmes d'information des ministères chargés de l'Éducation et de l'Enseignement supérieur.

³ Bien qu'anachronique, le terme « SYSCA » a été retenu ici parce qu'il correspond au nouveau nom du système d'information mis en place depuis l'année scolaire 2017/2018 pour remplir les fonctionnalités des Bases Élèves académiques de l'ancien système d'information « Scolarité ».

à ce stade, d'envisager de se passer d'un questionnement minimum sur les phases clés du parcours scolaire dans le corps du questionnaire de l'enquête Génération.

Le principal écueil rencontré tient à l'insuffisante disponibilité d'un INE valide pour les individus de la base de sondage de l'enquête Génération.

Le rapprochement entre l'enquête et les données administratives s'appuie en effet sur l'utilisation de l'INE. Disponible pour la DEPP dans ses fichiers, le Céreq s'efforce aussi de le recueillir de façon temporaire⁴ pour construire la base de sondage de son enquête afin de traiter les doublons et de repérer une partie des poursuites d'études. Pour rapprocher les résultats de l'enquête et les données administratives, la DEPP a joué le rôle de tiers de confiance en injectant un identifiant non signifiant du Céreq dans les extraits de ses tables. L'INE ne figure donc pas dans les fichiers du rapprochement expérimental.

Malheureusement, la disponibilité d'un INE valide est très variable selon les composantes de la base de sondage. Elle est excellente sur les champs « SYSCA » et « SIFA », alimentés par la DEPP, ainsi que sur le champ des universités publiques. Ensemble, ces composantes couvrent 70 % des répondants à l'enquête Génération 2017, mais toutes composantes confondues, 21 % des répondants n'ont pas d'INE valides. Or ces non-répondants sans INE valides alimentent de façon variable l'ensemble des niveaux de sortie et l'ensemble des catégories de plus haut diplôme atteint, qui constituent les deux variables d'entrée utilisées pour étudier le lien entre la formation initiale et l'insertion professionnelle. Ainsi, pour une approche par niveaux de sortie, les seuls niveaux très bien couverts (avec quasiment 100 % d'INE supposés valides) sont les sortants de collège, de terminales générales ou technologiques. Les sortants de CAP-BEP, de terminales professionnelles et de L3 et M2 sont aussi assez bien couverts, avec environ 90 % d'INE supposés valide. À l'opposé, les sortants de formation de niveau bac+2 à bac+4 dans les domaines de la santé et du social et les sortants d'école de commerce ou d'ingénieurs sont mal couverts. Cette variabilité se retrouve évidemment dans l'approche par le plus haut diplôme.

L'absence des informations relatives aux formations et examens relevant d'autres ministères, notamment le ministère de l'Agriculture, atténue le champ des personnes retrouvées au moins une fois dans les fichiers administratifs mobilisés et la capacité à suivre les parcours.

Pour les individus retrouvés au moins une fois, l'appariement avec les fichiers des apprenants semble fonctionner plutôt bien.

Le sexe et l'année de naissance figurant dans les fichiers des apprenants sont quasi toujours identiques à ceux de l'enquête Génération. Pour les répondants concernés, les départements de résidence en classe de sixième et au baccalauréat sont aussi largement concordants quand l'information est retrouvée dans les données administratives (l'information administrative est disponible pour 51 % des répondants concernés pour le lieu de résidence en sixième, et pour 65 % des répondants concernés pour le lieu de résidence au baccalauréat). Plus globalement, les parties de parcours scolaires que les données administratives permettent de documenter sont cohérentes avec les niveaux de sortie et les plus hauts diplômes atteints pour les répondants.

Dans le cas particulier du baccalauréat, examiné plus en détail, 34 % des répondants déclarant avoir le baccalauréat à l'enquête ne sont pas identifiés comme tels avec les données administratives mobilisées. Si l'on se limite aux répondants retrouvés au moins une fois dans ces données, cette proportion tombe à 7 %. Pour les 93 % repérés comme étant bachelier dans les deux sources, l'année du baccalauréat est identique dans 93 % des cas. La mention est identique dans 92 % des cas⁵. Concernant le type de baccalauréat, les informations sont cohérentes quand les données administratives ont enregistré un baccalauréat général ou professionnel. En revanche, s'il s'agit d'un baccalauréat technologique, l'enquête converge dans seulement 78 % des cas. Nous n'avons pu déterminer les raisons de cet écart ciblé.

⁴ L'INE n'est pas conservé par le Céreq à l'issue de la collecte. Le rapprochement expérimental a pu être réalisé parce qu'il a été anticipé en tenant compte du calendrier de production de l'enquête.

⁵ Quand il y a divergence sur la mention au bac, le cas où elle est meilleure dans l'enquête que dans les données administratives est six fois plus fréquents que le cas inverse, ce qui pourrait suggérer l'existence d'un léger biais déclaratif.

La profondeur historique retenue, qui couvre douze années scolaires consécutives, n'est pas suffisante pour couvrir l'ensemble du parcours dans le secondaire d'une partie des répondants. Pour les personnes retrouvées au moins une fois, quatre années supplémentaires permettraient d'atteindre un bon taux de couverture de l'ensemble du parcours scolaire dans le secondaire, quel que soit leur plus diplôme atteint, excepté pour les doctorats. Pour ces derniers, couvrir l'ensemble du secondaire nécessiterait une profondeur beaucoup plus grande. Dans leur cas, l'arbitrage entre le coût d'exploitation et l'apport pour les analyses peut se poser.

Comme dans toutes les données d'origine administrative, la qualité des informations proposées est variable. Le cas de la PCS des parents est ainsi l'occasion d'illustrer et de rappeler plusieurs limites posées par le recours aux données administratives à la place de données d'enquête :

- la dépendance de l'information collectée aux pratiques administratives des acteurs (PCS des personnes « responsables », plutôt que des « parents ») ;
- l'absence d'homogénéité entre acteurs dans la façon de collecter une information de même nature (PCS d'un seul responsable dans SIFA, de deux responsables dans SYSCA) ;
- une rigueur moindre dans la collecte d'informations non essentielles pour les opérations de gestion de l'acteur (les taux d'information non renseignée sont considérables dans SIFA) ;
- les risques de perturbations externes (absence d'information en 2008).

Il faut donc privilégier les informations au cœur des objectifs de gestion des systèmes d'information placés en amont des fichiers utilisés. Lorsque l'information est récupérable, les fichiers des apprenants permettent ainsi de connaître les classes successives dans le secondaire, de repérer les passages en SEGPA, les redoublements au collège ou au lycée, le passage par le secteur privé, l'inscription et les résultats en CAP, au baccalauréat, au BTS...

Si un appariement systématique devait être envisagé à l'avenir pour les prochaines enquêtes Génération, il ne faut pas sous-estimer la charge de travail qu'il nécessiterait. En particulier :

- l'utilisation des fichiers nécessite un temps d'appropriation, incluant des travaux exploratoires, pour bien comprendre l'origine et la nature des informations disponibles, ainsi que les hétérogénéités potentielles existantes entre les différentes sources agrégées. Ces travaux ne peuvent être faits une fois pour toutes. Ils seraient à renouveler si l'appariement était renouvelé ;
- pour que les informations issues des données administratives soient réellement exploitées, il ne faudrait mettre à disposition des chercheurs qu'une sélection d'informations choisies et synthétiques, à définir et à construire à partir de l'ensemble des fichiers annuels successifs⁶ ;
- les fichiers annuels comportent des individus en doublon. Il faut donc définir comment traiter ces doublons. Faut-il en tenir compte pour la construction des variables synthétiques choisies ou faut-il se limiter à privilégier une ligne par rapport aux autres ? Dans ce dernier cas, retient-on celle que le producteur des fichiers sources a choisie ou non ? Nous avons fait des choix en amont de ce travail expérimental, mais ils ne s'imposent pas forcément comme une évidence.
- Les fichiers disponibles peuvent enregistrer deux inscriptions à un examen pour la même année scolaire. Là encore, faut-il tenir compte de ces deux informations quand elles existent ou non ? Là aussi, nous avons fait un choix argumenté qui serait à rediscuter.
- Pour définir les classes successives suivies, une nomenclature agrégée doit être construite à partir de la nomenclature détaillée FORMATRANS, que la DEPP a élaborée pour pouvoir exploiter de façon transversale les informations de SYSCA et de SIFA. Ce travail ne peut pas être fait de façon définitive, une fois pour toutes, car la nomenclature de départ peut évoluer en fonction de l'évolution des politiques publiques. Ce travail nécessite aussi une bonne connaissance des marges du système éducatif, ainsi que des choix en partie subjectifs.

⁶ Par exemple, pour chaque classe, l'année, la région, le type d'établissement, ainsi que des indicateurs de redoublement au collège, au lycée, de passage par des classes SEGPA, etc. En cas de redoublement, il faudra définir si toutes les informations sont retenues ou si une année est privilégiée et une indicatrice de redoublement ajoutée. Pour les examens retenus, il faudra voir si toutes les inscriptions doivent être documentées, seulement les réussites aux examens ou seulement le fait d'avoir réussi au moins une fois tel ou tel examen. Corrélativement, il faudra réfléchir à la structure des tables : une seule table avec une multitudes de variables, une table de parcours, une table d'examen, etc.

En conclusion, l'appariement expérimental réalisé invite à conserver pour le moment dans le questionnaire de l'enquête Génération des questions clés sur le parcours scolaire, comme le redoublement au collège, l'orientation à l'issue de la troisième ou la réussite au baccalauréat, qu'il conviendrait de poser à tous, quel que soit leur niveau de sortie. Cette condition est indispensable pour conserver des informations garanties et homogènes pour tous, sans complexifier excessivement la phase de préparation de l'enquête et les filtres du questionnaire.

De façon complémentaire et non substituable, des extraits des fichiers des apprenants permettraient d'enrichir la connaissance des parcours des sortants ayant atteint un niveau pour lequel l'appariement est efficace en termes de taux de couverture : les sortants de collège, de terminales générales ou technologiques, de CAP-BEP et de terminales professionnelles. Pour eux, il serait nécessaire de définir les indicateurs qui pourraient être introduits de façon à ne proposer qu'un extrait synthétique et facilement exploitable des données de la DEPP, plutôt que des extraits bruts des fichiers annuels des apprenants.

Dans la perspective d'une prochaine enquête, il conviendrait de voir s'il est possible d'améliorer significativement la récupération de l'INE pour les segments mal couverts. Si le faire lors de la constitution de la base de sondage apparaît difficile, une alternative, plus simple techniquement, serait de s'assurer de la bonne identification des personnes lors de la phase de collecte de l'enquête en collectant l'essentiel de leurs informations d'état-civil. La disponibilité de ces informations rendrait possible leur identification par la DEPP au RNIE. Cette piste, possible techniquement, doit évidemment être examinée avec la DEPP, pour s'assurer de sa faisabilité juridique⁷, pour évaluer son coût et pour voir dans quelle mesure la collecte d'une partie seulement de l'état civil, permettrait d'atteindre une bonne couverture⁸.

Il conviendrait de voir également s'il est possible de disposer d'un fichier analogue aux fichiers des apprenants pour les formations délivrées par les établissements dépendant du ministère de l'Agriculture.

Quand les données de gestion permettront de connaître les grandes étapes du parcours scolaire de l'essentiel des sortants, quel que soit leur plus haut diplôme atteint, il sera possible de réduire le questionnaire de l'enquête. En particulier, même s'il sera nécessaire de continuer de poser des questions sur les classes suivies après la troisième, sur le type de baccalauréat et la classe post-baccalauréat, afin d'introduire des questions sur l'orientation à ces deux moments clés du parcours, il ne sera plus nécessaire de dérouler les questions destinées à définir la spécialité précise des classes d'inscription et diplômes passés ou les résultats aux examens.

⁷ Une organisation possible pourrait s'inspirer de celle retenue pour ce travail expérimental : le Céreq transmettrait une table de passage entre un identifiant « Céreq » non signifiant et l'INE quand il en dispose ou les informations d'état civil, sinon. Pour les répondants sans INE, une nouvelle étape interviendrait. Pour eux, la DEPP procéderait à une identification à son répertoire pour retrouver leur INE. La suite se déroulerait ensuite comme pour le rapprochement expérimental. La DEPP procéderait aux extractions des fichiers des apprenants en les restreignant aux observations associées aux INE des répondants ; elle injecterait l'identifiant « Céreq » dans ces tables, supprimerait les INE puis transmettrait les extraits ainsi construits au Céreq.

⁸ Par exemple, en se limitant à valider le nom de naissance, prénom, sexe, année et mois de naissance, département ou pays de naissance, sans collecter le jour précis de naissance, ni la commune de naissance.

1. Les données mobilisées et les objectifs du rapprochement expérimental

1.1 Une présentation succincte des données mobilisées

1.1.a) L'enquête Génération 2017 du Céreq

Depuis le début des années 1990, le dispositif d'enquêtes « Génération » permet au Centre d'études et de recherches sur les qualifications (Céreq) de **documenter régulièrement les premières années de vie active des jeunes après leur sortie de formation initiale**. Ces enquêtes interrogent de façon homogène tous les jeunes sortis du système éducatif, des jeunes arrêtant leurs études au niveau du collège sans aucun diplôme, jusqu'aux docteurs. Comme tous les enquêtés sont sortis de formation initiale au cours ou à l'issue d'une même année scolaire, il est facile de **comparer les conditions d'accès au marché du travail en fonction du niveau et de la nature de la formation initiale suivie** : niveau de diplôme, spécialités, passage par une filière professionnelle, l'apprentissage, etc. Jusqu'à présent, les enquêtes Génération sont la seule source labellisée « statistique publique » qui le permettent. Depuis le début des années 1990, huit cohortes de sortants ont été enquêtées⁹.

L'enquête Génération 2017 porte sur les jeunes sortis de formation initiale au cours ou à l'issue de l'année scolaire 2016/2017 (voir encadré 1). La passation du questionnaire, réalisée par internet et par téléphone, était initialement prévue au printemps 2020. La survenue de l'épidémie de Sras-Cov2 a obligé à repousser la collecte en raison du confinement général de la population décidé par le gouvernement du 17 mars au 10 mai 2020, pour lutter contre l'épidémie. La collecte a finalement eu lieu d'août 2020 à mars 2021. Les premiers résultats de l'enquête ont été publiés en mai 2022 (Coupié, Gaubert, Personnaz, 2022).

Le questionnaire de l'enquête aborde le parcours scolaire et les diplômes obtenus. Ce module permet d'identifier le plus haut diplôme obtenu et de recueillir d'autres informations sur la trajectoire scolaire, susceptibles de jouer sur le lien entre formation et emploi. Un calendrier mensuel d'activité collecte des informations sur la situation professionnelle mois par mois depuis la fin des études intervenue environ trois ans auparavant. D'autres modules permettent d'aborder des thématiques complémentaires, dans certains cas à la demande d'acteurs institutionnels. Pour l'enquête Génération 2017, des questions abordent ainsi les séjours à l'étranger, l'attractivité de la fonction publique, les moyens de financement des études, la multiactivité, la satisfaction dans l'emploi occupé à la date d'enquête, le sentiment de discrimination, l'exposition aux risques physiques et chimiques, ainsi que les conséquences sur la situation professionnelle du confinement décrété au printemps 2020 (Jugnot, Vignale, 2022).

Le module sur le parcours scolaire interroge notamment les jeunes sur des sujets pour lesquels les sources administratives disponibles auprès de la Direction de l'évaluation, de la prospective et de la performance¹⁰ (DEPP) disposent aussi d'informations sur leurs champs :

- le fait d'avoir redoublé ou non avant l'entrée en sixième,
- la classe de troisième suivie,
- l'orientation effective après la troisième,
- le fait d'avoir redoublé au collège,
- le fait d'avoir obtenu un baccalauréat, le type de bac, la spécialité et la mention obtenue,
- le fait de disposer d'un CAP, d'un BEP, d'une mention complémentaire ou d'un autre diplôme de niveau équivalent.

Dans l'enquête Génération, pour des raisons de coûts et de charge d'enquête, certaines de ces questions ne sont pas posées à tous les sortants, mais seulement aux sortants ayant un niveau d'études supérieur ou égal à la troisième, mais inférieur ou égal à bac+3. De leur côté, comme nous le verrons, les données de la DEPP ne couvrent pas l'ensemble du secondaire ni l'ensemble des examens. Il existe donc des trous de couverture qui peuvent, par exemple, perturber le repérage des redoublements ou de l'obtention de certains diplômes.

⁹ Le dispositif « Génération » a été présenté à la commission emploi du Conseil national de l'information statistique (Cnis), le 9 novembre 2017. Le diaporama de présentation et le compte-rendu de la réunion sont disponibles auprès du Cnis. <https://www.cnis.fr/evenements/emploi-qualification-et-revenus-du-travail-2017-2e-reunion-2/>

¹⁰ La DEPP est le service statistique ministériel, au sens de la loi n° 51-711 du 7 juin 1951, du ministère chargé de l'éducation nationale.

Encadré 1 • Le champ de l'enquête Génération 2017

La « Génération 2017 » porte sur les élèves et les étudiants considérés comme des « primo sortants » de formation initiale entre octobre 2016 et octobre 2017. Tous les niveaux et domaines de formation sont concernés.

Plus précisément, pour appartenir au champ de l'enquête, il faut :

- avoir été inscrit dans un établissement de formation en France, en métropole ou dans un département d'outre-mer, durant l'année scolaire 2016-2017 ;
- avoir quitté le système éducatif entre octobre 2016 et octobre 2017
- ne pas avoir interrompu ses études durant dix-sept mois¹¹ ou plus avant l'année scolaire 2016-2017, sauf pour raison de santé ;
- ne pas avoir repris ses études pendant les seize mois qui ont suivi l'entrée sur le marché du travail ;
- avoir trente-cinq ans ou moins en 2017.

Ces conditions sont cumulatives.

Pour connaître la trajectoire de l'ensemble des sortants des écoles dont ils ont la responsabilité, certains ministères demandent des extensions d'échantillon au Céreq, de manière à couvrir également les sortants qui ne sont pas primo-sortants. Dans le cas de l'enquête Génération 2017, des sortants « post-initiaux » ont ainsi été interrogés pour les établissements relevant du ministère chargé des Sports. Ces sortants ne sont pas pris en compte dans les exploitations réalisées par le Céreq. Ils ne sont donc pas pris en compte dans nos travaux exploratoires.

Enfin, dans l'enquête Génération 2017, les contrats de professionnalisation à visée diplômante sont traités comme les contrats d'apprentissage lorsqu'ils interviennent dans la continuité du parcours scolaire initial : ils sont considérés comme relevant de la formation initiale même s'il s'agit juridiquement de contrats de travail. Une personne repérée comme sortant de formation en 2016-2017 qui poursuit par un contrat d'apprentissage ou un contrat de professionnalisation en 2017-2018 est donc considérée en poursuite d'études, et donc hors champ. En revanche, les jeunes sortants de contrats d'apprentissage ou de contrat de professionnalisation au cours de l'année scolaire 2016-2017 sont dans le champ de l'enquête si ces contrats prolongent un parcours scolaire initial achevé antérieurement.

1.1.b. Les fichiers annuels des apprenants de la DEPP¹²

Depuis le début des années 2000, la **DEPP produit des fichiers anonymisés d'élèves pour la recherche et les études. Ils sont élaborés à partir de différentes sources administratives relevant du ministère chargé de l'Education nationale.** Ces fichiers de données individuelles, exhaustives dans leur champ, ne contiennent pas d'informations directement identifiantes, telles que les noms, prénoms et coordonnées. En revanche, ils disposent de l'INE, l'identifiant national des élèves et étudiants pour permettre l'articulation des différents fichiers entre eux afin de reconstituer des parcours, dans un cadre strict d'utilisation.

Le fichier de l'année n comprend :

- **tous les élèves inscrits un moment au cours de l'année scolaire $n/n+1$ dans les bases Élèves Académie (BEA)** à partir du système d'information « Scolarité », qui couvre l'ensemble des établissements du second degré publics ou privés sous contrat du ministère de l'Éducation nationale (dans la suite nous parlerons de « SYSCA » par abus de langage, puisque SYSCA est le nouveau système d'information mis en place en 2017-2018 pour remplacer Scolarité). Les élèves scolarisés

¹¹ L'enquête Génération 2017 est la première à utiliser le seuil de seize mois d'interruption. Auparavant, il ne fallait pas avoir interrompu ses études pendant douze mois consécutifs avant l'année scolaire de référence pour entrer dans le champ de l'enquête. Le changement de seuil résulte du développement des périodes de césure dans l'enseignement supérieur. Celles-ci permettent aux étudiants inscrits dans une formation initiale d'enseignement supérieur, de suspendre temporairement leurs études pendant un semestre ou deux semestres consécutifs « dans le but d'acquérir une expérience personnelle ou professionnelle, soit en autonomie, soit encadré dans un organisme d'accueil en France ou à l'étranger » (article D611-13 du code de l'Éducation). Selon l'établissement d'inscription et la formation suivie, cette période peut permettre d'acquérir des crédits ECT.

¹² La présentation des fichiers des apprenants s'appuie principalement sur le document de présentation produit par la DEPP, intitulé *Fichier des apprenants (dictionnaire des variables), rentrée 2003 à 2017* (32 pages).

dans le second degré dans les établissements relevant du ministère de l'Agriculture, dans les lycées militaires et dans la majorité des établissements privés hors contrat ne sont pas couverts par ces systèmes d'information¹³ ;

- **les élèves inscrits en apprentissage en janvier de l'année $n+1$** sont également inclus depuis 2008 à partir du système d'information SIFA – ils sont donc absents des fichiers annuels 2005, 2006 et 2007 ;
- enfin, des informations sur les **inscriptions et résultats à certains examens** issues des systèmes d'information développés pour les gérer, principalement le brevet, les baccalauréats, les CAP et BEP, ainsi que les BTS¹⁴. Dans tous les cas, les spécialités agricoles ne sont pas couvertes.

Pour les élèves inscrits au cours de l'année scolaire $n/n+1$ dans un établissement du champ « SYSCA » ou en janvier de l'année n pour le champ « SIFA », le fichier annuel dispose de caractéristiques de l'élève (sexe, date de naissance, département de naissance, nationalité agrégée, commune de résidence, PCS des adultes « responsables » et lien entre l'élève et ces adultes). Il propose des descripteurs de l'établissement de scolarisation (numéro d'établissement, département et académie, caractère public/privé sous contrat ou hors contrat, type d'établissement), ainsi que des informations associées à la scolarité de l'élève (classe d'inscription, formation suivie, régime scolaire, matières aux choix, date et motif de sortie, identifiant de la classe, boursier ou non). Dans le détail, **les informations disponibles dépendent des systèmes d'information source et des pratiques des gestionnaires**¹⁵. Elles évoluent dans le temps, de même que **certaines nomenclatures utilisées, en fonction des enrichissements décidés par la DEPP, mais aussi du fait d'évolutions des politiques publiques et des systèmes de gestion décidés en amont**¹⁶.

Pour les examens disponibles, le fichier annuel permet de repérer les élèves inscrits à l'examen, leur présence effective et les résultats obtenus (mention). Depuis 2013, deux examens peuvent être pris en compte dans le fichier annuel pour un élève donné¹⁷.

Comme nous l'avons vu, les fichiers annuels des apprenants sont produits par la DEPP à partir de données administratives issues de plusieurs systèmes de gestion. Chaque année, la DEPP effectue donc un rapprochement entre ces fichiers. Dans un premier temps, elle constitue un fichier des élèves inscrits dans les différentes classes et formations couverts par son champ à partir d'extractions des systèmes d'information SYSCA et SIFA. Elle repère ensuite les cas d'inscriptions multiples et, dans ce cas, elle définit l'inscription prioritaire à prendre en compte pour produire des statistiques sans doubles comptes, sans supprimer pour autant les autres observations (voir encadré 2). Dans un second temps, la DEPP procède à l'enrichissement du fichier des apprenants avec les informations issues des fichiers des examens. **L'ensemble de ces rapprochements s'effectue en s'appuyant sur l'INE.**

¹³ Selon la documentation de la DEPP, « 3 % des élèves du 2nd degré et environ 15 % des élèves du post-bac sont exclus du fichier des apprenants et donc de l'analyse des parcours ».

¹⁴ Sur la période que nous étudions, le système d'information OCEAN (Organisation des Concours et Examens Académiques et Nationaux) donne des informations pour le baccalauréat, le CAP, le BEP, le BP, les mentions complémentaires de niveau V et IV ainsi que certains diplômes supérieurs (BTS et diplômes comptables). Il est maintenant remplacé par le système d'information CYCLADES.

¹⁵ En particulier, le département de naissance est absent des fichiers SIFA jusqu'en 2016 inclus. La nationalité est détaillée pour les élèves des fichiers SYSCA, mais pas pour ceux des fichiers SIFA (pour eux, distinction « Union européenne » / « Autre »). La PCS est collectée pour deux personnes « responsables » dans les fichiers SYSCA et une seule personne dans les fichiers SIFA (de plus, dans SIFA, la non-réponse est proche de 50% et dépasse 60% pour les apprentis du supérieur). La situation d'emploi des personnes responsables était collectée dans les fichiers SYSCA, mais pas dans les fichiers SIFA ; elle n'est plus collectée pour la voie scolaire à partir de 2017. L'identifiant de la classe, le fait d'être boursier et le régime scolaire (situation par rapport à l'hébergement ou la demi-pension) ne sont connus que dans les fichiers SYSCA. Pour les élèves quittant leur établissement en cours d'année scolaire, la date de sortie et le motif ne sont disponibles que dans les fichiers SYSCA.

¹⁶ Par exemple, pour les nomenclatures détaillées des formations suivies.

¹⁷ Les cas où deux inscriptions à un examen sont enregistrées dans les fichiers des apprenants sont rares. Par exemple, parmi les 173 025 individus échantillonnés pour l'enquête Génération 2017 retrouvés dans le fichier des apprenants de 2016, 86 369 disposent au moins d'une inscription connue dans le fichier des apprenants. Parmi eux, seulement 1 021 ont une deuxième inscription, soit 1,2 %. Il s'agit essentiellement d'élèves inscrits au baccalauréat professionnel, également inscrits à un examen de CAP ou de BEP et plus rarement d'élèves inscrits en CAP, inscrits en même temps dans un autre CAP ou un BEP.

Encadré 2 • Le traitement des inscriptions multiple une année donnée

L'existence d'inscriptions multiples est documentée par la DEPP. Elle peut notamment résulter d'un changement d'académie de scolarisation en cours d'année (« doublons internes » à SYSCA) ; d'une présence simultanée dans SYSCA et SIFA résultant d'un champ partiellement commun aux deux systèmes d'information ou d'un passage d'un système à l'autre en cours d'année scolaire.

La variable TYPE_DOUBLON permet de catégoriser les différentes situations possibles :

1	« <i>INE uniques</i> »	
2	« <i>Vrais doublons Sco</i> »	INE présents plusieurs fois dans la source SYSCA et possédant les mêmes caractéristiques individuelles pour le sexe, la date et le département de naissance.
3	« <i>Faux doublons Sco</i> »	INE présents plusieurs fois dans la source SYSCA, mais possédant au moins une différence sur les trois caractéristiques individuelles retenues <i>supra</i> .
4	« <i>Vrais doublons Sco/Sifa</i> »	INE présents deux fois, une fois dans la source SYSCA et une fois dans la source SIFA (avec les mêmes caractéristiques individuelles, sexe et date naissance).
5	« <i>Faux doublons Sco/Sifa</i> »	INE présents deux fois, une fois dans la source SYSCA et une fois dans la source SIFA, mais possédant au moins une caractéristique individuelle différente (sexe et date naissance).
6	« <i>Doublons Sco et Sco/Sifa</i> »	INE présents trois fois dans le fichier, deux fois dans la source SYSCA et une fois dans la source SIFA.

Pour traiter ces cas de présences multiples, la DEPP détermine un enregistrement prioritaire de façon à pouvoir effectuer des tabulations sans doubles comptes. Cet enregistrement prioritaire est identifié par la variable indicatrice FLAG_DOUBLON, qui vaut alors 1 :

<i>INE unique</i> (Type_doublon=1)	Flag_doublon=1.
<i>Vrai doublon Sco</i> (Type_doublon=2)	Un enregistrement est priorisé (flag_doublon = 1). Pour les autres, flag_doublon=0. Avant 2011, l'enregistrement priorisé était choisi aléatoirement. À partir de 2011, si pour un même INE, le motif de sortie est renseigné pour un enregistrement et non renseigné pour un autre enregistrement alors ce dernier enregistrement est priorisé. Dans les autres cas, le choix est effectué par tirage aléatoire.
<i>Vrai doublon Sco/Sifa</i> (Type_doublon=4, certains Type_doublon=6)	L'enregistrement issu de Sifa est priorisé (flag_doublon = 1). Celui ou ceux issus de SYSCA sont écartés (flag_doublon = 0).
<i>Faux doublons</i> (Type_doublon=3 et 5, les autres Type_doublon=6)	Aucun enregistrement n'est priorisé. Pour tous, flag_doublon=0. L'INE n'est donc pas pris en compte dans les tabulations sans doubles comptes.

Source : DEPP, document de présentation des fichiers FAERE.

Avant la mise en place de SIFA en 2017, environ 15 % des apprentis n'avaient pas d'INE renseigné. Dans ce cas, les élèves concernés se voient attribuer un INE fictif dans le fichier des apprenants. La proportion d'INE fictifs était variable d'une académie à l'autre et d'une année sur l'autre¹⁸.

Dans les fichiers de gestion des examens, la proportion d'INE renseignés est très variable selon le type d'examens et l'année considérée. Si la proportion d'identifiants renseignés approche les 100 % pour les candidats au bac, elle n'atteignait par exemple que 17 % pour les inscrits au CAP en 2009. Pour ces observations sans INE, la DEPP procède à des traitements spécifiques, qui mobilisent différentes

¹⁸ De 2013 à 2016, six académies avaient un taux d'INE fictif supérieur à 20%, dont Paris et Versailles.

caractéristiques individuelles pour réattribuer les informations disponibles dans les fichiers des examens aux élèves potentiellement concernés présents dans les fichiers des inscrits (voir encadré 3).

Encadré 3 • Attribuer un examen au bon élève, même en l'absence de l'INE

La récupération des informations sur les examens nécessite d'apparier le fichier des inscrits avec chacun des fichiers des examens disponibles, soit sept fichiers. **Parmi les inscrits, seuls les enregistrements priorisés¹⁹ sont pris en compte.**

Pour faciliter l'opération, la DEPP définit un vivier de candidats potentiels pour chaque examen. Il s'agit d'isoler les formations pour lesquelles il est raisonnable de chercher à enrichir les données de scolarité par d'éventuels résultats aux examens. Par exemple, le vivier du baccalauréat comprend les inscrits dans des classes de terminales (générales, technologiques ou professionnelles), de première année de BTS ou DTS, de mise à niveau BTS, ainsi que les classes pour l'inclusion scolaire (UPI-ULIOS), les modules d'accueil en lycée de niveau IV et les FCII de niveau IV²⁰.

Dans une première phase, l'appariement est fait à partir de l'INE. L'ensemble des élèves appartenant au vivier de l'examen concerné est apparié au fichier de l'examen correspondant. Pour que l'enrichissement soit validé pour un individu donné, la concordance de l'INE ne suffit pas. Il faut aussi une cohérence entre les caractéristiques individuelles présentes dans le fichier des apprenants et celles du fichier de l'examen (les caractéristiques individuelles retenues sont le sexe et la date de naissance). C'est pour le brevet et le baccalauréat que cette phase est la plus « efficace », avec un appariement de près de 90 % des individus présents dans le fichier de l'examen.

Dans une deuxième phase, l'appariement est fait sur profil. Pour les élèves inscrits dans une classe ou une formation débouchant sur un examen au cours ou à la fin de l'année scolaire²¹, une variable de profil « strict » est construite en compilant les informations suivantes : sexe, date et département de naissance, établissement fréquenté, diplôme et spécialité préparés. Deux nuances sont à apporter à la construction de ce profil « strict » :

1° pour les élèves en apprentissage, le département de naissance n'est pas utilisé pour construire la variable de profil avant 2017 puisque l'information n'était pas présente dans les bases SIFA ;

2° la réforme de la voie professionnelle a compliqué l'exercice en permettant aux apprenants inscrits en première professionnelle en vue d'un baccalauréat professionnel de s'inscrire à un diplôme intermédiaire, le CAP ou le BEP. À partir de la session 2009 du CAP et du BEP, il a donc fallu en tenir compte et construire une table de passage entre les spécialités du baccalauréat professionnel et celles des diplômes intermédiaires correspondants.

Quand la variable de profil « strict » permet d'attribuer un examen à un unique apprenant, l'observation du fichier de l'examen lui est attribuée.

Enfin, une variable de profil « dégradé » est constituée, qui ne prend pas en compte la formation suivie, afin de voir si les examens restants peuvent être attribués à des élèves inscrits dans des classes ou formations relevant du vivier de l'examen considéré sans en être l'année terminale naturelle.

Source : DEPP, document de présentation des fichiers FAERE.

Pour nos travaux expérimentaux, douze fichiers annuels des apprenants successifs ont été mobilisés pour couvrir les années 2005 à 2016. Il s'agissait de récupérer des informations sur le parcours dans l'enseignement secondaire jusqu'au baccalauréat pour la plupart des niveaux de sortie couverts par l'enquête Génération. Plus précisément, la profondeur historique des données administratives demandée a été déterminée de façon à couvrir les années dans l'enseignement secondaire des jeunes de la Génération 2017 jusqu'aux sortants de niveau bac+5 dès lors qu'ils ont connu une scolarité linéaire. Comme nous le verrons,

¹⁹ Flag_doublon=1.

²⁰ La composition des viviers est présentée dans le document méthodologique de la DEPP présentant les fichiers FAERE.

²¹ Classe de troisième pour le diplôme national du brevet, classe terminale pour le baccalauréat, année terminale de CAP (BEP) pour le CAP (BEP).

l'expérimentation montre que ce choix n'est pas optimal. Une profondeur historique plus importante aurait été préférable.

Trois facteurs déterminent la capacité à reconstituer des parcours à partir de ces données d'origine administrative :

- **la présence de l'élève dans un établissement ou une formation relevant du champ de ces données.** Celles-ci couvrent l'essentiel, mais pas la totalité de l'enseignement secondaire. Une absence d'un élève attendu une année donnée peut donc résulter d'une scolarité à l'étranger, d'une déscolarisation ou d'une scolarité sur le territoire national dans un établissement ou une formation « hors champ » (formations agricoles, établissements privés hors contrat, scolarité à domicile...) ;
- **la qualité du renseignement de l'INE par les services gestionnaires** dans leurs fichiers. Comme nous l'avons vu plus haut, celle-ci n'était pas excellente pour les données de SIFA jusqu'en 2016, générant des INE « fictifs ». L'absence d'INE peut créer des trous dans les parcours alors même que l'information est disponible dans les fichiers. Elle dégrade aussi la capacité à récupérer des informations relatives aux examens ;
- **la stabilité de l'INE dans le temps.** L'INE a été mis en place comme identifiant de gestion unique destiné à suivre les élèves durant toute leur scolarité. En pratique, des débats politiques ont conduit à réinitialiser l'INE à l'entrée dans le secondaire, tandis que les modes de gestion ont pu entraîner des changements d'INE en cours de scolarité, notamment en cas de changement d'académie dans le secondaire et dans l'enseignement supérieur. Dans ces cas, seuls des morceaux de parcours peuvent être reconstitués, avec chacun des INE successifs associés à l'élève.

À l'avenir, les deux dernières limites devraient disparaître ou, au moins, s'amoinrir. Le ministère de l'Éducation nationale a confié à la DEPP la mise en place d'un répertoire national des identifiants élèves, étudiants et apprentis (RNIE) et des nouveaux systèmes d'information ont été mis en place pour s'articuler avec ce répertoire²². Ces évolutions, déjà mises en œuvre, ne concernent pas la période couverte par notre appariement expérimental, mais faciliteront un éventuel renouvellement de l'exercice, pour la partie des parcours concernée.

1.2. Les objectifs du rapprochement expérimental

Depuis quelques années, la facilitation et le développement des utilisations des données administratives à des fins de recherche et de production statistique créent de nouvelles opportunités (Jugnot, 2021), notamment pour les enquêtes qui, comme les enquêtes Génération, sont labellisées « Statistique publique » au sens de loi n° 51-711 du 7 juin 1951 sur l'obligation, la coordination et le secret en matière de statistiques. En particulier, la mise en œuvre d'un identifiant national des élèves unique et certifié, ainsi que les investissements réalisés par la DEPP sur ses données administratives, contribuent à ces nouvelles opportunités. Ensemble, ils devraient conduire progressivement à mieux suivre les parcours scolaires à partir des données administratives issues des processus de gestion. De ce fait, un rapprochement de ces données avec l'enquête Génération pourrait permettre de réduire les coûts et la charge d'enquêtes en raccourcissant la durée d'interrogation. Certaines informations permettraient aussi d'enrichir les informations collectées. Il faudrait aussi voir si les informations disponibles dans ces sources administratives pourraient également améliorer le traitement de la non-réponse globale, sur tout ou partie du champ de l'enquête. Dans cette perspective, un appariement expérimental a été réalisé entre l'enquête Génération 2017 et les fichiers annuels des apprenants de la DEPP pour examiner les potentialités de ces nouvelles opportunités et mieux connaître ces données. Ce document de travail en présente les résultats.

Un exercice analogue est en cours pour la connaissance des parcours de l'enseignement supérieur avec des données produites par le service statistique ministériel de l'Enseignement supérieur. Des sources administratives permettent aussi d'apporter des informations sur l'accès au marché du travail. Un premier appariement a ainsi été réalisé avec les données de Pôle Emploi (Jugnot, 2022).

²² Pour plus de détail, voir (Jugnot, 2021).

Ayant une vocation d'information générale et d'études, ces appariements ne sont pas exhaustifs : ils ne portent que sur un échantillon représentatif de sortants du système éducatif. Ils n'ont pas vocation à être systématiques puisque le dispositif « Génération » porte sur des générations de sortants espacés dans le temps – désormais tous les quatre ans. Ce rythme permet à la fois de documenter les liens structurels entre la formation initiale et les conditions d'accès à l'emploi, qui n'évoluent que lentement dans le temps, ainsi que les inflexions conjoncturelles induites par la conjoncture économique.

De façon plus précise, nous souhaitons répondre à plusieurs questions :

- Quelle est la proportion du champ de l'enquête Génération pour lequel des informations sont retrouvées dans les données administratives proposées par la DEPP ?
- Dans quelle mesure, ces données, grâce à leur profondeur historique, permettent de fournir des indicateurs de parcours pour les répondants de l'enquête (redoublements, réussites à certains examens, orientation après la troisième...) ?
- Les informations issues des données administratives et les informations recueillies dans l'enquête Génération sont-elles cohérentes ?
- Dans quelle mesure, ces données administratives peuvent aider au traitement de la non-réponse ?

Les réponses dépendent en partie de la capacité à appairer les individus échantillonnés dans l'enquête Génération avec les individus présents dans les données administratives. Sans entrer dès maintenant dans la présentation détaillée de l'appariement, celui-ci s'appuie sur l'identifiant national des élèves et étudiants, l'INE. Sa disponibilité dans l'enquête Génération est variable. Elle dépend de la source mobilisée pour alimenter la base de sondage de l'enquête. De ce fait, la qualité de l'appariement varie selon les segments de cette base de sondage et, corrélativement, selon le plus haut diplôme atteint et le niveau de sortie. La plupart des exploitations réalisées sont détaillées selon ces trois informations : composantes de la base de sondage, plus haut diplôme atteint²³ et niveau de sortie²⁴. Les tableaux et illustrations présentés dans le corps du texte sont détaillés, quand il y a lieu, selon les deux premières dimensions. Les mêmes résultats détaillés selon le niveau de sortie sont regroupés dans l'annexe 2.

2. L'organisation pratique de l'appariement

Le rapprochement expérimental de l'enquête Génération 2017 avec les fichiers des apprenants de la DEPP est encadré par une convention signée par le Céreq et la DEPP le 3 septembre 2021²⁵. Il a été déclaré le 29 octobre 2021 au registre des activités de traitement du Céreq, conformément à l'article 30 du règlement général sur la protection des données (RGPD).

2.1. Le rôle clef de l'INE

L'Identifiant national des élèves, étudiants et apprentis, a vu le jour en 1995 sous le nom de « numéro de matricule national ». Il était alors destiné aux établissements publics de l'enseignement secondaire relevant du ministère de l'Éducation nationale, dans le cadre de la mise en place du système d'information Scolarité, pour gérer les inscriptions des élèves dans les établissements et le pilotage des effectifs au niveau des académies et au niveau national (SYSCA a remplacé récemment ce système d'information). D'emblée, l'objectif affiché par le ministère de l'Éducation nationale était de disposer d'un identifiant unique, destiné à suivre chaque élève durant toute sa scolarité, de façon à faciliter les différentes opérations de gestion : les inscriptions des élèves dans les établissements, l'orientation, les inscriptions aux examens, etc. Son champ d'utilisation s'est progressivement élargi et ce déploiement continue, tant quant aux établissements couverts qu'en termes d'outils de gestion le mobilisant. L'INE joue ainsi, dans la sphère éducative, le rôle qu'a le numéro d'identification au répertoire de l'Insee (le « numéro de sécurité sociale ») dans la sphère sociale – l'usage du NIR avait été un temps envisagé par l'Éducation nationale, mais la Commission nationale informatique et

²³ Le plus haut diplôme obtenu est une information construite par le Céreq à partir du niveau de sortie et des autres informations collectées sur le parcours scolaire. C'est un référentiel propre au Céreq.

²⁴ Le niveau de sortie est une information construite par le Céreq à partir de la classe dans laquelle l'élève est inscrit au cours ou à l'issue de l'année scolaire de sortie (2016-2017). C'est un référentiel propre au Céreq (la base de diffusion de l'enquête Génération propose aussi l'information selon le référentiel de l'Insee).

²⁵ Convention n°2021/09/DEEVA.

libertés l'a explicitement refusé pour éviter que tous les fichiers administratifs n'utilisent le même identifiant (Jugnot, 2021).

Présent dans les différents systèmes d'information de l'Éducation nationale, l'INE joue un rôle central dans la constitution des fichiers annuels des apprenants de la DEPP (voir partie 1.1.b). Dans les fichiers mis à disposition des chercheurs et du Céreq, il n'est pas directement présent, mais une version cryptée est disponible, pour permettre les appariements entre les fichiers annuels successifs. Dans l'enquête Génération, l'INE est aussi disponible pour une partie importante de l'échantillon car le Céreq le collecte autant que possible pour constituer la base de sondage de l'enquête (encadré 4). Le Céreq ne peut le conserver que pour la phase de production de l'enquête. Même si le Céreq ne dispose pas l'INE pour tout son échantillon, il été retenu pour procéder à l'appariement expérimental entre l'enquête Génération 2017 et les fichiers des apprenants en l'absence d'autres variables d'identification dans les fichiers des apprenants, notamment des patronymes et adresses. De plus, l'appariement sur la base d'un identifiant de gestion est préférable à un appariement sur des informations moins normées.

Comme l'INE n'est pas conservé dans les fichiers de résultats des enquêtes Génération, il ne devait pas figurer dans les fichiers de travail de l'appariement expérimental. L'organisation pratique du rapprochement entre les données du Céreq et celles de la DEPP a dû tenir compte de cette obligation. Concrètement, l'équipe chargée de la production des enquêtes Génération au Céreq a produit une table de passage entre les identifiants non signifiants du dispositif Génération et les INE, qu'elle détenait pour la production de l'enquête dans un répertoire non accessible aux chargés d'études. Cette table de passage a été transmise à la DEPP. Le service habilité a ensuite procédé au cryptage des INE de cette table puis la DEPP a extrait les observations associées à ces INE cryptés des fichiers annuels des apprenants 2005 à 2016. Elle a enfin remplacé l'INE crypté par l'identifiant non signifiant du Céreq et mis ces fichiers à disposition du Céreq. Les exploitations du rapprochement expérimental ont donc été effectuées sur des fichiers ne contenant ni l'INE, ni l'INE crypté.

Encadré 4 • Construire la base de sondage de l'enquête Génération 2017

Il n'existe pas de fichier centralisé contenant l'ensemble des élèves et étudiants inscrits une année scolaire donnée et moins encore, de fichiers permettant d'identifier les seuls primo-sortants qui constituent le champ d'intérêt des enquêtes Génération. Avant chaque nouvelle « génération », le Céreq doit donc construire sa base de sondage de façon *ad hoc*, dans le respect de la loi Informatique et Libertés et de la loi n° 51-711 du 7 juin 1951 sur l'obligation, la coordination et le secret en matière de statistiques.

Pour l'enquête Génération 2017, le Céreq a ainsi collecté des listes d'élèves et d'étudiants inscrits au cours des années scolaires 2016-2017 et 2017-2018 dans les établissements du second degré et de l'enseignement supérieur, en s'efforçant de récupérer l'INE, quand il était disponible. Disposer de deux années scolaires consécutives permet de supprimer une partie des poursuites d'études de la base de sondage. Disposer de l'INE facilite ce travail, notamment en cas de changement d'établissement d'une année sur l'autre. L'INE permet aussi de repérer des personnes inscrites dans plusieurs établissements au cours de l'année scolaire de référence pour éviter les doubles comptes dans la base de sondage. Ces travaux préalables permettent ainsi de réduire la taille de la base de sondage pour cibler le mieux possible le strict champ de l'enquête. Cependant, l'appartenance à ce champ, notamment le fait que l'élève ou l'étudiant est réellement primo-sortant, ne peut être validée définitivement que lors de la passation du questionnaire. Il commence par un module dédié à cela.

Plus précisément, les informations concernant certains établissements ou formations sont récupérées de façon centralisée. C'est le cas pour :

- l'essentiel des établissements du second degré, grâce à la DEPP, à partir du Système d'information statistique consolidé académique (SYSCA) ;
- l'essentiel des formations par apprentissage, grâce à la DEPP, à partir du Système d'information de la formation des apprentis (SIFA) ;
- les lycées agricoles ou autres établissements relevant du ministère de l'Agriculture, grâce à sa Direction générale de l'enseignement et de la recherche ;
- les écoles d'architecture et d'arts couverts par la plateforme TAIGA du ministère en charge de la Culture ;
- les formations « jeunesse et sports » couvertes par l'outil de gestion Formations et diplômes de l'animation et du sport (FORÔMES), de la Direction de la jeunesse de l'éducation populaire et de la vie associative (DJEVPA) du ministère de l'Éducation nationale ;
- les inscrits dans une convention industrielle de formation par la recherche (CIFRE), grâce à l'Association nationale de la recherche et de la technologie (ANRT) ;
- les contrats de professionnalisation, grâce à la Direction de l'animation de la recherche, des études et des statistiques (DARES) ;

Pour les établissements qui ne sont pas couverts par ces collectes centralisés, le Céreq procède à une collecte d'informations en s'adressant directement à eux. C'est notamment le cas pour les universités, les écoles d'ingénieurs, les écoles de commerce, les écoles de la santé et du social. La liste des établissements à enquêter est établie à l'aide de la Base centrale des établissements gérée par le ministère en charge de l'Éducation.

La base de sondage des enquêtes Génération agrège ainsi plusieurs composantes, pour lesquelles la récupération de l'INE est plus ou moins facile et plus ou moins exhaustive. Par exemple, l'INE est un identifiant de gestion pour SYSCA et SIFA alors qu'il est absent du système d'information des contrats de professionnalisation. Dans le cas des collectes directes auprès des établissements, la disponibilité de l'INE dépend de son usage dans les systèmes de gestion, de la facilité à l'extraire avec les listes d'inscrits et de l'acceptation de l'établissement à le transmettre.

2.2. Un individu sur cinq de l'échantillon n'a pas d'INE valide

L'échantillon final de l'enquête Génération 2017 compte 303 573 individus, dont 266 063 appartiennent à l'échantillon principal et 37 510 à l'échantillon de réserve qu'il a fallu engager pour atteindre un nombre de répondants suffisant, soit 25 164 questionnaires dans le champ des primo-sortants étudiés par le Céreq et 1 207 questionnaires supplémentaires de « post-initiaux » appartenant à une extension d'échantillon sur les formations « jeunesse et sports »²⁶.

Pour certains individus, le Céreq disposait de plusieurs INE dans sa base de sondage. Dans ce cas, le Céreq a dupliqué les lignes dans la table de passage de façon à transmettre à la DEPP tous les couples (IDENT, INE) possibles. La moitié des individus identifiés par leur IDENT, soit 151 385 individus, sont présents deux fois dans la table de passage et 117 le sont trois fois (tableau 1).

Au total, la table de passage transmise à la DEPP comprend 455 192 observations et trois variables : l'INE, l'identifiant non signifiant du Céreq (IDENT), ainsi qu'une indicatrice indiquant si l'individu a répondu ou non à l'enquête.

La plupart des individus qui disposent de plusieurs INE sont issus des composantes SYSCA et SIFA de la base de sondage car la DEPP a transmis au Céreq les deux variables dont elle disposait pour l'INE : d'une part, l'INE dit « RNIE », correspondant à l'INE enregistré au nouveau répertoire national alors en cours de déploiement pour assurer l'unicité de l'identifiant ; d'autre part, l'INE « BEA », qui était attribué au niveau académique dans le cadre du mode de gestion existant auparavant. Ces doublons représentent 98 % des individus présents deux fois dans la table de passage. En les mettant de côté, seuls 1 % des individus de l'échantillon final du Céreq sont associés à deux INE différents ou plus. Tous appartiennent au segment « universités » de la base de sondage.

Tableau 1 • Nombre de lignes dans la table de passage associés à un IDENT donné selon la composante de la base de sondage

Composante de la base de sondage	Nombre de lignes			Total
	1	2	3	
Sysca (second degré)	5 801	104 485		110 286
Sifa (apprentissage)	18 846	43 937		62 783
Universités	46 781	2 963	117	49 861
Autres établissements	28 882			28 882
Agriculture	8 865			8 865
DREES	4 225			4 225
Sport	11 536			11 536
Culture	1 107			1 107
DGAFP	1 517			1 517
Cifre	890			890
Contrats de prof.	23 621			23 621
Total	152 071	151 385	117	303 573

Champ : échantillon (répondants et non répondants).

Source : table de passage transmise par le Céreq à la DEPP.

Si un individu peut avoir plusieurs INE, la réciproque n'est normalement pas possible. Un INE donné ne doit correspondre qu'à un seul individu. La manière d'attribuer l'INE, tant dans le champ de la DEPP que dans celui de l'Enseignement supérieur, a été organisée de façon à le garantir. Par ailleurs, si l'individu n'ayant qu'un INE a été inscrit simultanément, ou successivement, dans deux établissements au cours de l'année

²⁶ Le ratio entre le nombre de réponses exploitables et l'échantillon final peut sembler bas par rapport à des enquêtes plus classiques de la Statistique publique, mais, à la différence de ces enquêtes, le Céreq ne dispose pas d'une base de sondage « idéale » couvrant uniquement le champ désiré, avec des coordonnées disponibles et à jour.

scolaire 2016-2017, cette double inscription est normalement repérée par le Céreq lors du traitement des doublons effectué pour la constitution de la base de sondage. Il devrait donc y avoir une unicité des valeurs prises par l'INE dans la base de sondage, donc dans la table de passage transmise à la DEPP. Une vérification a été effectuée de façon sommaire pour le confirmer.

Une tabulation simple montre que 405 120 observations de la table de passage, soit 89 %, sont associées à des valeurs uniques pour l'INE. Les 11 % restantes correspondent à des valeurs d'INE manifestement invalides :

- dans 49 723 cas, l'INE est vide ;
- dans 40 cas, l'INE vaut « 00000000000 » ;
- dans 28 cas, l'INE vaut « 0 » ;
- dans 9 cas, l'INE vaut « 2 » ;
- dans 7 cas, l'INE vaut « 2016/2017 » ;
- dans 5 cas, l'INE vaut « 1861558 »
- dans 5 autres cas, il vaut « NA » ;
- dans 4 cas, l'INE vaut « / » ;
- dans 4 autres cas, il vaut « INCONNU »

Nous avons donc cherché à quantifier la proportion d'INE manifestement invalides dans la table de passage transmise à la DEPP. À cet effet, après avoir examiné quelques extraits de la table ainsi que les valeurs prises par la variable INE lorsqu'elles sont présentes de façon multiple, nous avons considéré comme manifestement invalides :

- les valeurs d'une longueur inférieure à 11 positions (11^{ème} position vide, plus exactement),
- les valeurs entièrement numériques,
- les valeurs commençant par une lettre,
- les valeurs comprenant une lettre, mais commençant par « 99999 ».

Cette approche conduit à considérer qu'au moins **11,3 % des observations de la table de passage ont manifestement un INE invalide**. Les individus issus des composantes SYSCA, SIFA et Universités de la base de sondage ne sont pas concernés. À l'opposé, les individus issus des composantes Cifre, Contrats de professionnalisation et Sports ont tous des INE invalides puisque ces sources ne fournissent pas l'INE. Pour les établissements relevant des autres composantes, la situation est intermédiaire (tableau 2). Dans leur cas, il existe sans doute des situations isolées de méconnaissance ou de mauvaise saisie de l'INE pour certains de leurs inscrits, mais l'ampleur des informations manquantes conduit plutôt à supposer que certains établissements n'ont systématiquement pas transmis l'INE au Céreq, parce qu'il n'était pas présent ou facilement extractible de leur système de gestion, ou parce qu'ils ont refusé de le faire.

Tableau 2 • Nombre et proportion d'INE valides dans la table de passage, par composante de la base de sondage

Composante de la base de sondage	% d'INE invalides	INE invalide	dont valeur vide	INE supposé valide	Ensemble
Sysca (second degré)	0%			214 771	214 771
Sifa (apprentissage)	0%			106 720	106 720
Universités	0%	6		53 052	53 058
Autres établissements	35%	10 246	9 512	18 636	28 882
Agriculture	0%			8 865	8 865
DREES	82%	3 457	3 338	768	4 225
Sport	100%	11 536	11 536		11 536
Culture	69%	762	650	345	1 107
DGAFFP	61%	926	176	591	1 517
Cifre	100%	890	890		890
Contrats de prof.	100%	23 621	23 621		23 621
Total	11%	51 444	49 723	403 748	455 192

Champ : échantillon (répondants et non répondants).

Source : table de passage transmise par le Céreq à la DEPP.

Si l'on se limite aux 26 383 personnes²⁷ repérées comme répondantes à l'enquête dans la table de passage (y compris les post-initiaux du Sport), 14 % appartiennent à des composantes ne disposant pas de l'INE (Cifre, Contrats de professionnalisation, formations « jeunesses et Sports). **Toutes composantes confondues, 21 % des répondants de l'enquête Génération n'ont pas d'INE valides.**

Tableau 3 • Nombre de répondants associés à des INE supposés valides, par composante de la base de sondage

Composante de la base de sondage	Effectifs	Répartition	Nombre d'INE invalides	% d'INE invalides
Sysca (second degré)	7 493	28%	0	0%
Sifa (apprentissage)	3 082	12%	0	0%
Universités	6 984	26%	0	0%
Autres établissements	3 054	12%	873	29%
Agriculture	810	3%	0	0%
DREES	722	3%	559	77%
Sport	1 998	8%	1 998	100%
Culture	147	1%	83	56%
DGAFP	267	1%	96	36%
Cifre	67	0%	67	100%
Contrats de prof.	1 759	7%	1 759	100%
Total	26 383	100%	5 435	21%

Champ : répondants (y compris post-initiaux du Sport).

Source : table de passage transmise par le Céreq à la DEPP.

On sait donc d'emblée qu'il ne sera pas possible de récupérer d'informations dans les fichiers de la DEPP pour au moins un répondant sur cinq. *A contrario*, les trois composantes de la base de sondage pour lesquelles l'INE est bien connu (SYSCA, SIFA, Universités) couvrent ensemble 67 % des répondants.

L'INE n'étant pas disponible dans les fichiers du rapprochement expérimental, il n'est pas possible d'évaluer directement la qualité de la variable INE selon le plus haut diplôme ou par niveau de sortie. Il est cependant possible de regarder la distribution des répondants entre les composantes de la base de sondage pour lesquelles l'INE est totalement disponible, celles pour lesquelles il ne l'est que très partiellement et celles pour lesquelles il ne l'est pas du tout (tableau 4).

²⁷ Personnes identifiées par leur IDENT, identifiant non significatif du Céreq.

Tableau 4 • Répartition des répondants à Génération 2017 par plus haut diplôme entre composantes de la base de sondage, selon la qualité de l'INE

	Issus d'une composante de la base de sondage...			Nbre de répondants total	Issus d'une composante de la base de sondage...		
	Avec l'INE	Avec un INE partiel	Sans l'INE		Avec l'INE	Avec un INE partiel	Sans l'INE
01-Non diplômés	1 715	55	9	1 779	96%	3%	1%
02-CAP-BEP-MC-Divers Niv.3	2 381	111	39	2 531	94%	4%	2%
03-Bac pro-BT-Divers Niv.4	3 708	702	545	4 955	75%	14%	11%
04-Bac technologique	801	124	26	951	84%	13%	3%
05-Bac général	780	162	17	959	81%	17%	2%
06-Bac+2 sauf santé/social	2 126	386	436	2 948	72%	13%	15%
07-Bac+2 à +4 santé/social	89	529	5	623	14%	85%	1%
08-Licence professionnelle	1 026	106	331	1 463	70%	7%	23%
09-Autre bac+3 et bac+4	976	407	266	1 649	59%	25%	16%
10-Bac+5 sauf écoles	2 917	867	647	4 431	66%	20%	15%
11-Ecoles (Ingé,commerce)	680	1 436	234	2 350	29%	61%	10%
12-Doctorat	354	110	61	525	67%	21%	12%
Ensemble	17 553	4 995	2 616	25 164	70%	20%	10%

Champ : répondants (hors post-initiaux).

Note : la pondération associée aux répondants de l'enquête Génération n'est pas utilisée ici.

Source : Céreq-Depp. Rapprochement de l'enquête Génération 2017 à trois ans et des fichiers des apprenants FAERE.

Un bémol complémentaire doit être apporté à ce tableau : les élèves inscrits dans les établissements relevant du ministère de l'Agriculture sont repérés comme issus d'une composante disposant d'un INE valide, mais ils sont hors du champ des fichiers des apprenants de la DEPP. Les informations sur les parcours les concernant n'y figurent donc pas. Plus de 9 % des répondants ayant comme plus haut diplôme, un bac professionnel, un BT ou un autre diplôme de niveau 4 appartiennent à cette composante de la base de sondage. Ils sont 6,5 % pour les bacs+2 hors santé/social et autour de 3 % pour les baccalauréats technologiques, les CAP et BEP et les non-diplômés.

2.3. Traitement des rares cas d'individus de l'échantillon présents plusieurs fois la même année dans les fichiers des apprenants

Le Céreq a exploité les extraits de douze fichiers annuels des apprenants successifs, couvrant les individus de l'échantillon de l'enquête Génération 2017 pour les années 2005 à 2016. Dans ces fichiers annuels, le nombre d'observations varie de 32 097 pour 2006 à 173 025 pour 2016 (tableaux 5a et 5b). Les personnes présentes en doublons, au sens de présence de plusieurs observations pour un même identifiant IDENT, sont rares. Dans un fichier annuel donné, la quasi-totalité des individus ne sont présents qu'une seule fois, de 97,2 % à 99,8 % selon les années. Au total, le nombre d'individus pour lesquels la DEPP a retrouvé des informations oscillent entre 32 032 individus dans le fichier 2006 et 169 973 individus dans le fichier 2016. En rapportant ces nombres aux 303 573 individus de la table de passage, la proportion d'individus retrouvés une année donnée oscille entre 11 % et 56 %. De façon logique, les taux les plus faibles concernent les années les plus anciennes : au fur et à mesure que l'on remonte dans le temps, les rapprochements possibles entre les deux sources se concentrent sur les sortants de la Génération 2017 qui ont eu la scolarité la plus longue.

Corrélativement, selon les années, de 0,2 % à 2,8 % des individus de l'échantillon de l'enquête Génération sont présents plusieurs fois une même année dans les fichiers de la DEPP. Les taux les plus élevés correspondent aux années les plus récentes car la majorité des doublons correspond à une présence simultanée dans les deux systèmes d'informations mobilisés, SIFA et SYSCA (« SCO »). Or l'essentiel de l'apprentissage est associé à des parcours scolaires relativement courts. Pour cette même raison, la proportion des doublons associés à une double inscription dans le système d'information SYSCA augmente au fur et à mesure que l'on remonte dans le temps, jusqu'à devenir majoritaire.

Comme nous l'avons vu dans la partie 1.1.b, la présence de doublons dans les fichiers annuels des apprenants n'est pas forcément une anomalie et ces doublons font l'objet d'un traitement spécifique lors de la création des fichiers des apprenants afin de déterminer l'observation à privilégier. Dans plus de 99 % des cas, les doublons sont des doublons stricts : l'individu (IDENT) est présent avec deux inscriptions dans le fichier des apprenants et pas davantage. Dans les très rares autres cas, ils sont présents avec trois observations.

La plupart des doublons repérés semble correspondre à des cas de « vrais » doublons au sens de la DEPP, puisque dans la quasi-totalité des cas, la variable flag_doublon prend la valeur 1 pour une observation et la valeur 0 pour l'autre. Dans ce cas, comme on l'a vu, la DEPP privilégie l'inscription en apprentissage sur l'inscription dans la voie scolaire lorsque le doublon résulte d'une présence simultanée dans SYSCA et SIFA. En cas de présence de doublon dans SYSCA, il peut y avoir en revanche une part de choix aléatoire dans l'inscription privilégiée.

Pour opérer le rapprochement avec Génération, il faut définir la façon de traiter ces cas de présences multiples, même si le sujet est mineur du fait de son ampleur limitée. Retient-on uniquement les observations priorisées par la DEPP (Flag_doublon=1) ? À l'opposé, exploite-t-on toutes les informations disponibles ? Pour ce travail exploratoire, compte tenu de la faible proportion de doublons, nous avons fait un choix intermédiaire : nous ne conservons qu'une seule ligne par individu (IDENT) en donnant la priorité à l'enregistrement associé à Flag_doublon=1 quand il y en a un. Dans les quelques cas où il n'y a aucun enregistrement priorisé par la DEPP et dans les cas, plus rares encore, où plusieurs observations sont priorisées, nous ne conservons qu'une observation, choisie aléatoirement. **Ce choix devrait être réexaminé si le rapprochement était renouvelé.**

Quelques informations jugées importantes issues des observations en doublons non retenues ont toutefois été conservées afin de permettre quelques vérifications sur leur apport possible²⁸.

²⁸ Variables suffixées « _L2 » pour la 2^e ligne et « _L3 » pour la troisième ligne dans le programme SAS.

**Tableau 5a • Les extraits des fichiers annuels des apprenants transmis au Céreq
(années 2011 à 2016)**

	Fichier 2016	Fichier 2015	Fichier 2014	Fichier 2013	Fichier 2012	Fichier 2011
Nombre total d'observations	173 025	158 810	150 653	153 172	149 758	147 718
dont source "SCO" (en %)	61.65	75.18	88.95	94.43	97.03	98.59
source "SIFA" (en %)	38.35	24.82	11.05	5.57	2.97	1.41
Avec flag_doublon=1 (en%)	98.15	97.21	98.31	98.77	99.09	99.32
Avec flag_doublon=0 (en%)	1.85	2.79	1.69	1.23	0.91	0.68
Nombre total d'IDENT différents	169 973	154 511	148 246	151 371	148 468	146 773
dont présents avec 1 observation	166 928	150 222	145 843	149 580	147 183	145 831
présents avec 2 observations	3 038	4 279	2 399	1 781	1 280	939
présents avec 3 observations	7	10	4	10	5	3
Cas des IDENT présents 2 fois						
2 obs. issues de "SCO" (en %)	11.92	12.36	25.84	41.44	51.56	70.71
1 issue de "SCO" et 1 de SIFA (en %)	88.02	87.64	74.16	58.56	48.44	29.29
2 issues de "SIFA" (en %)	0.07	0	0	0	0	0
2 obs. avec flag_doublon=1	0.07	0.02	0	0	0	0
2 avec flag_doublon=0	3.52	2.90	5.42	4.49	5.16	5.86
1 avec flag_doublon=1 et 1 avec flag_doublon=0	96.41	97.08	94.58	95.51	94.84	94.14

Champ : échantillon (répondants et non répondants).

Source : DEPP-Extraits des fichiers des apprenants 2011 à 2016.

**Tableau 5b • Les extraits des fichiers annuels des apprenants transmis au Céreq
(années 2005 à 2010)**

	Fichier 2010	Fichier 2009	Fichier 2008	Fichier 2007	Fichier 2006	Fichier 2005
Nombre total d'observations	133 267	101 278	76 198	56 368	41 421	32 097
Dont source "SCO" (en %)	99.24	99.44	99.57	100.00	100.00	100.00
source "SIFA" (en %)	0.76	0.56	0.43	Sans objet	Sans objet	Sans objet
Avec flag_doublon=1 (en%)	99.51	99.60	99.64	99.72	99.68	99.78
Avec flag_doublon=0 (en%)	0.49	0.40	0.36	0.28	0.32	0.22
Nombre total d'IDENT différents	132 650	100 887	75 939	56 212	41 296	32 032
dont présents avec 1 observation	132 038	100 497	75 681	56 059	41 172	31 967
présents avec 2 observations	607	389	257	150	123	65
présents avec 3 observations	5	1	1	3	1	0
Cas des IDENT présents 2 fois						
2 obs. issues de "SCO" (en %)	77.92	82.26	85.60	100.00	100.00	100.00
1 issue de "SCO" et 1 de SIFA (en %)	22.08	17.74	14.40	Sans objet	Sans objet	Sans objet
2 issues de "SIFA" (en %)	0	0	0	Sans objet	Sans objet	Sans objet
2 obs. avec flag_doublon=1	0.33	0.51	1.17	1.33	1.63	1.54
2 avec flag_doublon=0	5.44	5.40	5.84	2.00	8.13	10.77
1 avec flag_doublon=1 et 1 avec flag_doublon=0	94.23	94.09	93.00	96.67	90.24	87.69

Champ : échantillon (répondants et non répondants).

Source : DEPP-Extraits des fichiers des apprenants 2005 à 2010.

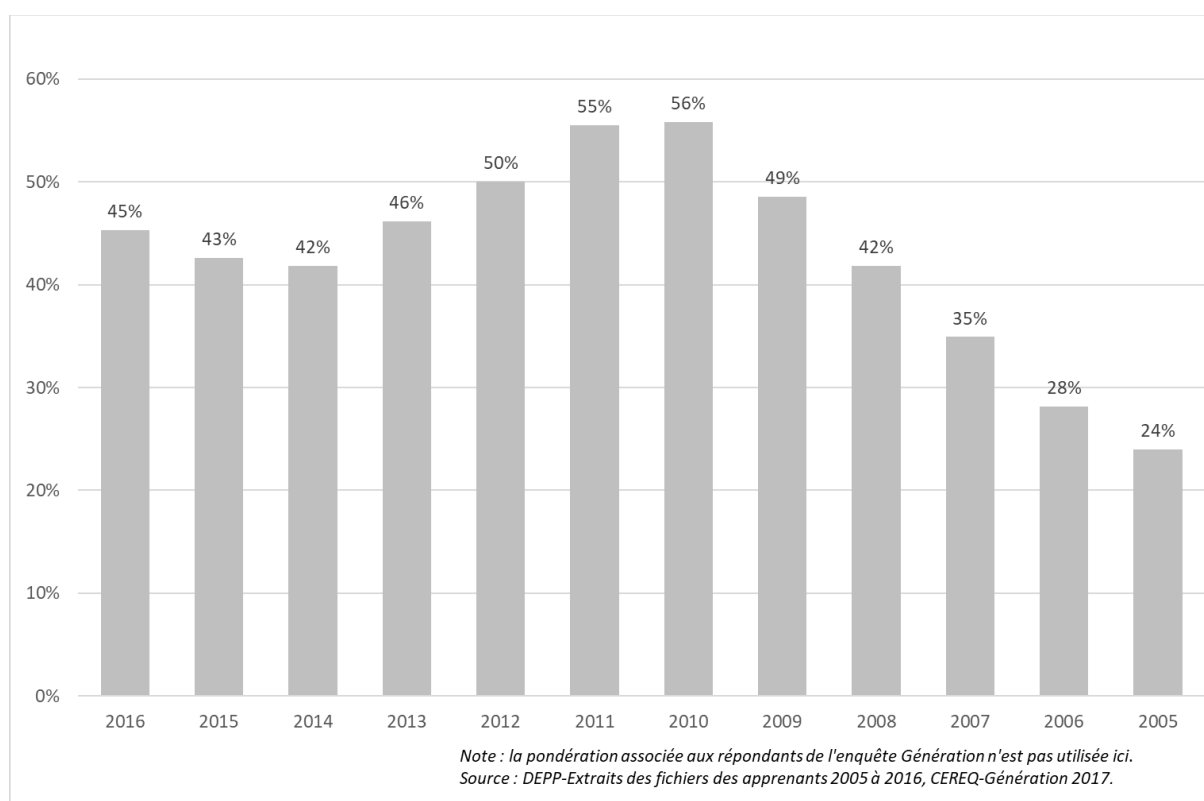
3. Quels répondants sont retrouvés dans les fichiers des apprenants ?

Cette partie présente quelques résultats de l'appariement expérimental du fichier de diffusion de l'enquête Génération 2017 avec les extraits des fichiers des apprenants 2005 à 2016. Il s'agit ici de voir quels répondants à l'enquête sont retrouvés dans les fichiers administratifs retravaillés par la DEPP, selon la composante de la base de sondage, le plus haut diplôme atteint et le niveau de sortie. Les publications du Céreq utilisant de façon privilégiée le plus haut diplôme atteint, celui-ci est privilégié pour les illustrations retenues dans le corps du texte. Celles relatives au niveau de sortie sont présentées dans l'annexe 2.

L'appariement est réalisé à partir des extraits des fichiers annuels des apprenants retravaillés, de façon à ne conserver qu'une observation par individu ayant répondu à l'enquête Génération (IDENT), comme indiqué à la fin de la partie précédente.

Les premières exploitations ont été réalisées avec la troisième, puis la quatrième version provisoire de la base de diffusion de Génération 2017, disponibles respectivement mi-novembre et fin novembre 2021. Elles ont été actualisées au cours de l'été 2022 avec la version définitive de la base et portent sur les 25 164 répondants figurant dans cette base. L'extension d'échantillon sur les « post-initiaux » du Sport n'est donc pas prise en compte.

Graphique 1 • Proportion de répondants à Génération 2017 retrouvés dans le fichier des apprenants de l'année considérée



Selon les années, entre 24 % et 56 % des répondants à l'enquête Génération 2017 sont retrouvés dans les fichiers annuels des apprenants. **Ces taux globaux ne suffisent pas pour juger de la qualité de l'appariement** car les fichiers de la DEPP couvrent surtout des établissements du second degré. La probabilité de retrouver des jeunes de la Génération 2017 une année donnée dépend donc de leur niveau de sortie. Les répondants ayant les plus bas niveaux de sortie doivent se retrouver dans les fichiers les plus récents, tandis que les répondants avec les plus hauts niveaux doivent se retrouver dans les fichiers les plus anciens. Le taux global synthétise cette combinaison de trajectoires, qui explique aussi que les taux d'appariement les plus

élevés ne sont pas observés pour les fichiers les plus récents : ils sont atteints pour les fichiers des apprenants 2010 et 2011, soit cinq et six ans avant l'année scolaire de sortie de la Génération 2017. Le taux le plus bas correspond au fichier annuel le plus ancien, douze ans avant l'année de sortie. Son niveau est loin d'être marginal, ce qui confirme **l'intérêt de recourir aux informations administratives sur une profondeur historique assez longue et sans doute davantage que celle utilisée pour ce rapprochement expérimental.**

Pour les répondants de l'enquête Génération retrouvés dans les fichiers annuels des apprenants, il est possible de comparer deux informations personnelles : l'année de naissance et le sexe. Globalement, ces informations sont cohérentes entre les deux sources puisqu'elles sont identiques à 99 % (tableau 6). Les rares cas de divergences n'ont pas été examinés et n'ont pas été supprimés de la base d'études pour les exploitations réalisées ensuite.

Tableau 6 • Cohérence de la date de naissance et du sexe entre les sources

	2016	2015	2014	2013	2012	2011	2010	2009	2008	2007	2006	2005
Même année de naissance et même sexe	11 272 98,9%	10 602 98,8%	10 424 98,9%	11 488 98,8%	12 435 98,8%	13 794 98,8%	13 877 98,8%	12 083 98,8%	10 420 98,9%	8 696 99,0%	7 027 99,1%	5 983 99,1%
Même année de naissance, mais sexe différent	51	53	43	52	60	76	77	67	50	39	27	20
Même sexe, mais année de naissance différente	77	72	68	79	90	93	88	76	64	50	36	33
Année de naissance et sexe différents	1	2	1	3	1	2	3	3	2	3	2	2

Champ : répondants (hors post-initiaux).

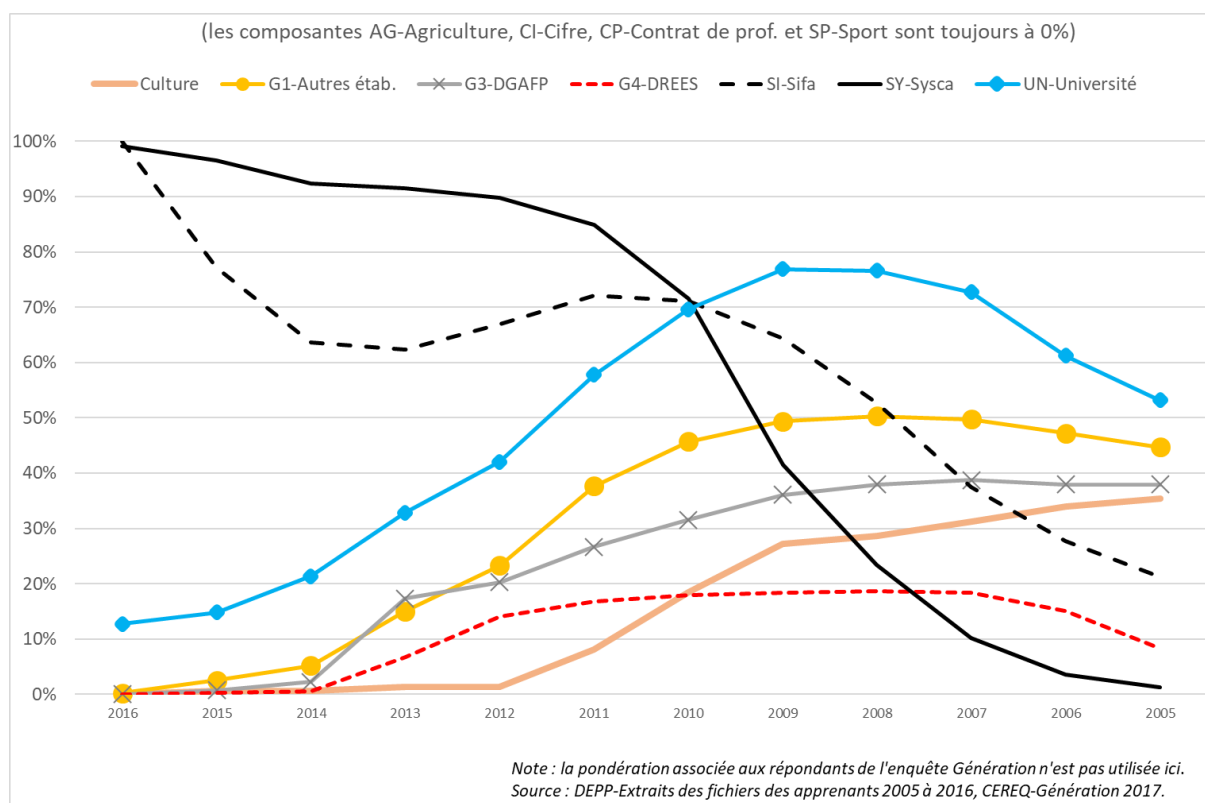
Note : la pondération associée aux répondants de l'enquête Génération n'est pas utilisée ici.

Source : Céreq-Depp. Rapprochement de l'enquête Génération 2017 à trois ans et des fichiers des apprenants FAERE.

L'examen des taux globaux d'appariement n'est pas suffisant pour évaluer la qualité du rapprochement des deux sources de données car il faut tenir compte de la capacité à récupérer l'INE, variable selon la composante de la base de sondage de l'enquête Génération. Le graphique 2 permet une analyse plus détaillée.

- Dans les champs SIFA et SYSCA, 100 % des répondants sont retrouvés dans le fichier annuel 2016 de la DEPP. Ce résultat est attendu puisque c'est la DEPP qui a fourni au Céreq les informations nécessaires, dont l'INE, pour alimenter cette composante de la base de sondage. Les taux d'appariements sont faibles pour les années les plus anciennes puisque ces fichiers couvrent surtout des jeunes de la Génération 2017 sortis au niveau du baccalauréat ou avant.
- Dans le champ des Universités, pour lequel l'INE est aussi largement disponible, le taux d'appariement est faible pour les années récentes, mais atteint près de 80 % en 2008-2009. Ce profil est aussi cohérent : la plupart des sortants des universités ne relèvent de l'Éducation nationale que jusqu'au baccalauréat, obtenu le plus souvent plusieurs années avant leur sortie de l'enseignement supérieur. Ce profil suggère aussi que le transfert de l'INE du secondaire vers l'enseignement supérieur s'effectue généralement bien pour les universités, puisque l'INE qu'elle ont transmis au Céreq permet de retrouver les étudiants dans les fichiers de l'Éducation nationale, notamment pour leurs parcours dans le secondaire, comme nous le verrons plus loin.
- Dans les champs couverts par la Culture, la DREES, la DGAFP et les autres établissements relevant de l'enseignement supérieur, le profil est analogue à celui du champ des universités, mais plus aplatis. Cette forme reflète les difficultés de récupération de l'INE dans la base de sondage pour ces champs.
- Pour l'Agriculture, l'INE récupéré par le Céreq est *a priori* valide. Il est disponible dans la base de sondage, mais les écoles relevant du ministère de l'Agriculture ne sont dans pas le champ des fichiers des apprenants.
- Pour le Sport, les contrats de professionnalisation et les CIFRE, l'INE n'est pas disponible dans la base de sondage.

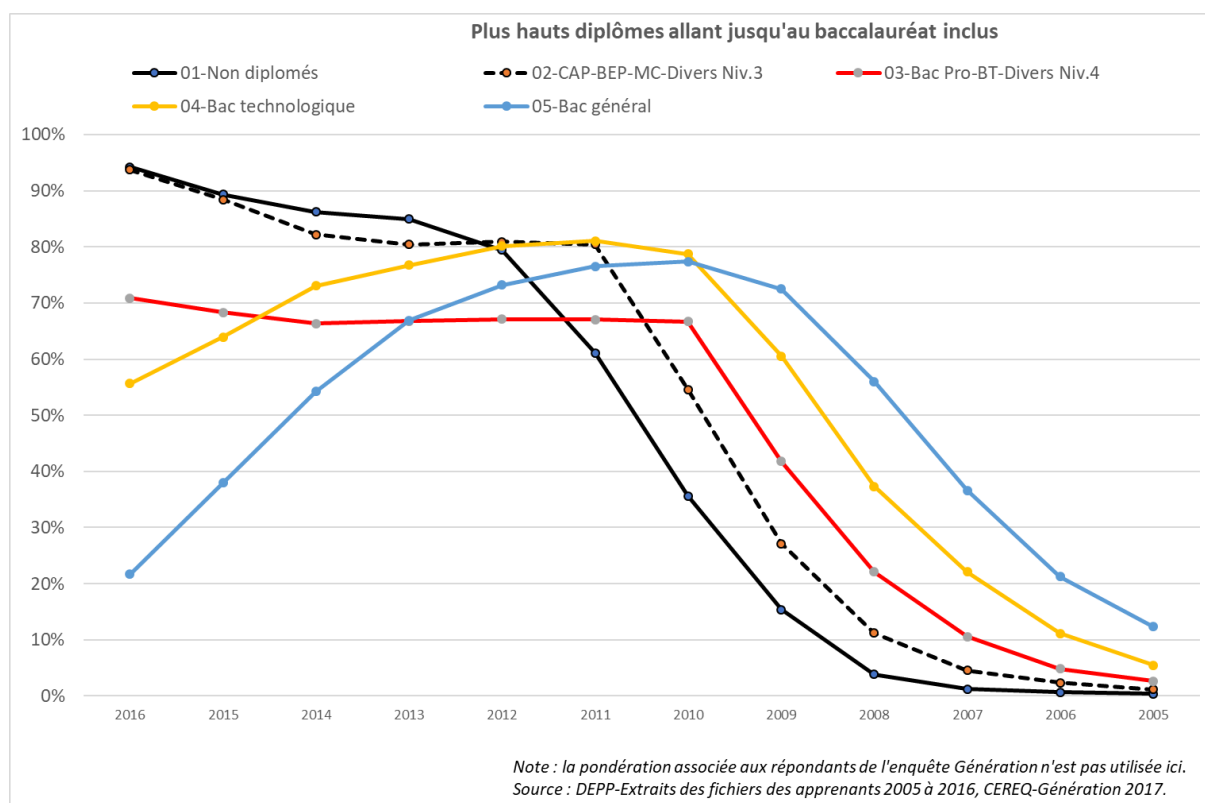
Graphique 2 • Proportion de répondants à Génération 2017 retrouvés dans le fichier FAERE de l'année considérée, selon la composante de la base de sondage d'appartenance



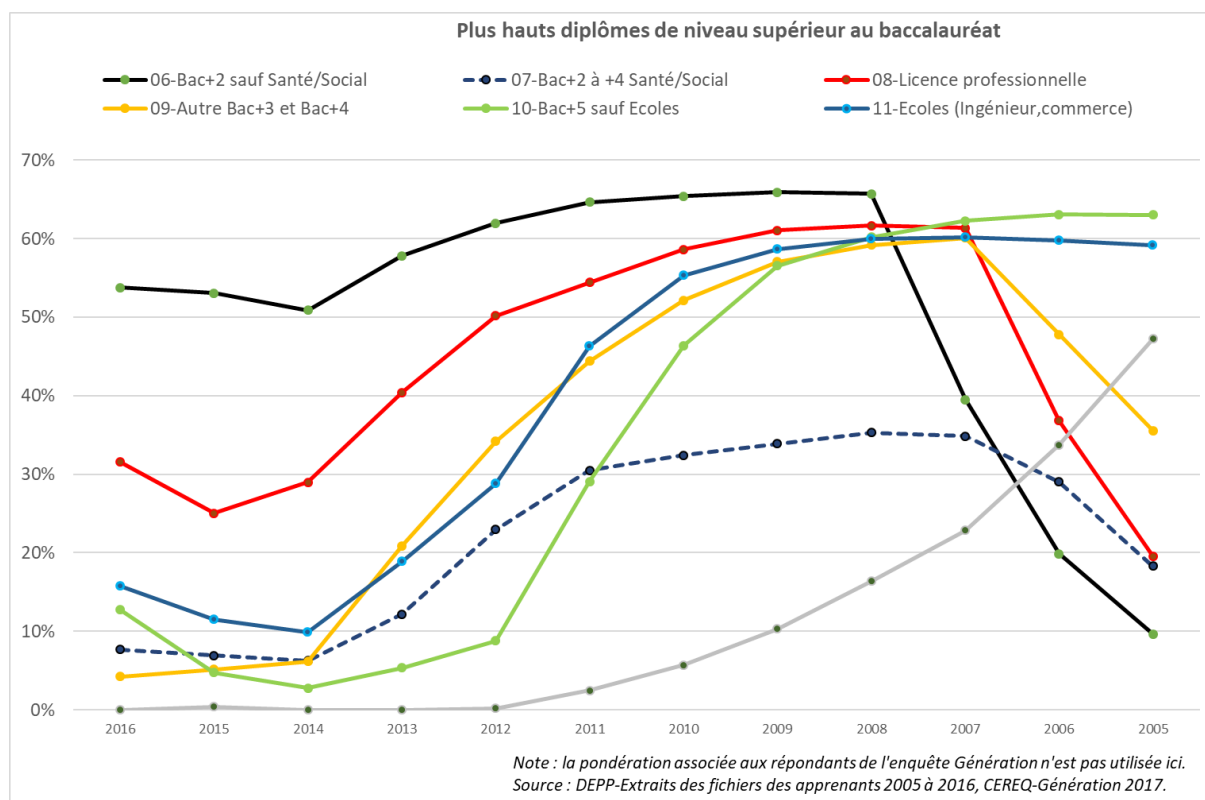
Les résultats des taux d'appariement selon le plus haut diplôme atteint traduisent les mêmes constats (graphique 3a et 3b). En particulier :

- pour les non-diplômés et pour les jeunes dont le CAP ou le BEP est le plus haut diplôme, l'appariement est très bon sur les années récentes et la tendance à les retrouver baisse ensuite, puisque les fichiers des apprenants ne couvrent pas les parcours avant le collège ;
- pour les jeunes dont le baccalauréat est le plus haut diplôme, le pic d'appariement n'est pas immédiat parce qu'une partie n'atteint ce niveau diplôme qu'après avoir tenté des études supérieures sans les réussir. Une partie d'entre eux relève par ailleurs de composantes de la base de sondage pour lesquelles l'INE est peu ou pas disponible ;
- pour les diplômés du supérieur, le constat est analogue. Plus le plus haut diplôme est élevé, plus le pic d'appariement est ancien car la sortie du secondaire est en moyenne plus ancienne. Par ailleurs, une partie importante des répondants concernés relève aussi de composantes où l'INE est peu ou pas disponible, ce qui diminue le maximum atteint par le taux annuel d'appariement.

Graphique 3a • Proportion de répondants à Génération 2017 retrouvés dans le fichier FAERE de l'année considérée, selon le plus haut diplôme atteint (jusqu'au baccalauréat)

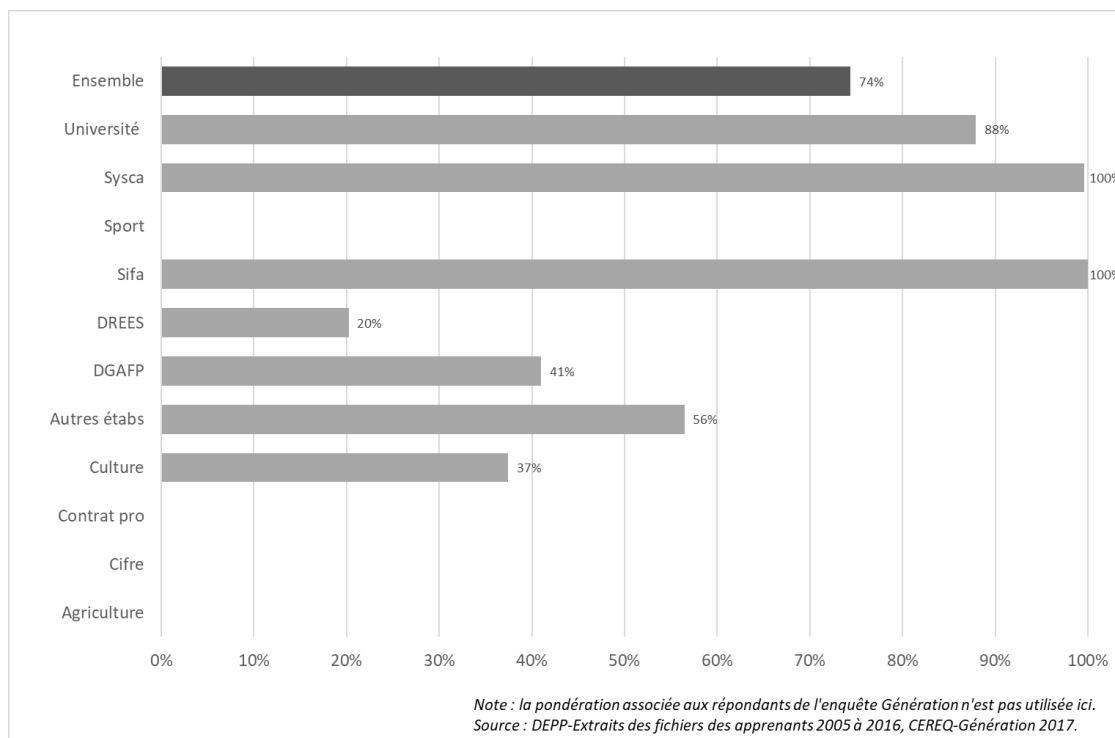


Graphique 3b • Proportion de répondants à Génération 2017 retrouvés dans le fichier FAERE de l'année considérée, selon le plus haut diplôme atteint (au-delà du baccalauréat)



Si l'on prend en compte l'ensemble des douze fichiers de la DEPP mobilisés, **75 % des répondants à l'enquête Génération 2017 sont retrouvés au moins une fois dans les fichiers des apprenants**. Tous les répondants associés aux composantes SYSCA et SIFA sont retrouvés au moins une fois, comme presque 90 % des répondants associés à la composante Universités. Pour les autres composantes, les taux sont nettement moins importants, voire nuls (graphique 4).

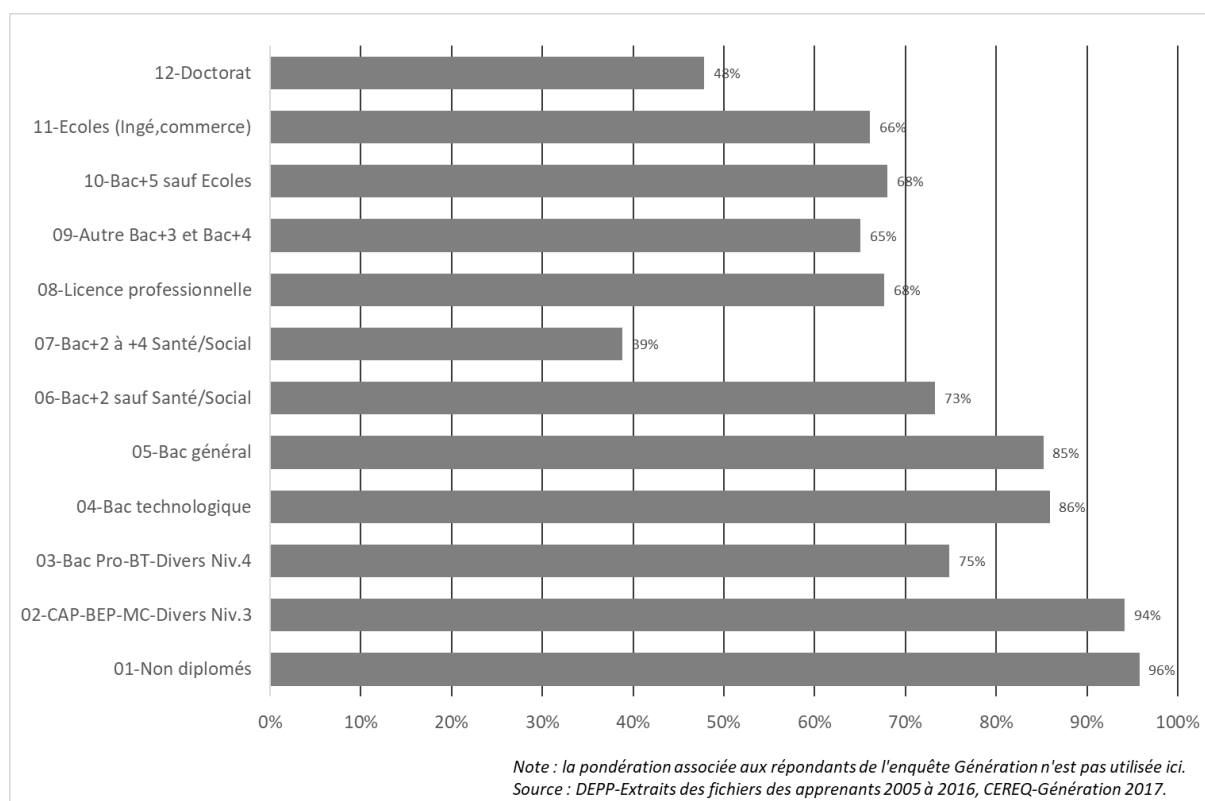
Graphique 4 • Proportion de répondants retrouvés au moins une fois parmi les fichiers des apprenants 2005 à 2016, selon la composante de la base de sondage d'appartenance



Ces différences sur la capacité à retrouver au moins une fois les répondants selon la composante de la base de sondage se répercutent sur la capacité à retrouver au moins une fois les répondants selon leur plus haut diplôme atteint (graphique 5) :

- la proportion de répondants retrouvés au moins une fois est très bonne pour les non diplômés et les jeunes ayant au mieux un niveau CAP-BEP, autour de 95 % ;
- cette proportion reste très élevée, mais de dix points inférieurs, au niveau du baccalauréat général et du baccalauréat technologique ;
- les baccalauréats professionnels et les bacs+2 hors santé/social viennent ensuite, avec encore environ dix points de moins, soit près d'un quart de répondants jamais retrouvés ;
- pour les autres niveaux de diplôme du supérieur, au moins un tiers des répondants ne sont jamais retrouvés.

Graphique 5 • Proportion de répondants retrouvés au moins une fois parmi les fichiers des apprenants 2005 à 2016, selon le plus haut diplôme atteint



Quand les répondants sont retrouvés au moins une fois dans les données administratives, ils le sont généralement plusieurs années et cela, quel que soit le niveau de sortie (tableau 7). Ils sont ainsi le plus souvent retrouvés :

- 5 à 8 fois pour les non-diplômés ;
- 6 à 8 fois pour les CAP-BEP ;
- 7 à 9 fois pour les bacheliers ;
- 7 à 10 fois pour les bacs+2 hors santé/social et les licences professionnelles ;
- 6 à 8 fois pour les bacs+3 et bacs+4 ;
- 5 à 7 fois pour les bacs+5 hors écoles ;
- 6 à 9 fois pour les écoles d'ingénieurs et de commerce ;
- 1 à 4 fois pour ceux qui sortent avec une thèse.

Ce constat rassure sur la capacité à reconstituer une grande partie des parcours à partir des données administratives pour les trois quarts des jeunes de la Génération 2017 qui sont retrouvés au moins une fois dans les fichiers des apprenants. L'examen de la première année observée (tableau 7), de la dernière année observée (tableau 8) et du nombre d'absences des fichiers des apprenants entre la première et la dernière année de présence (tableau 9) vont dans le même sens. De façon attendue, la première année d'observation est, en moyenne, d'autant plus ancienne que le plus haut diplôme atteint est élevé. Il en est de même pour la dernière année d'observation.

Ceux qui sortent du système éducatif avec une thèse font exception, la profondeur historique retenue n'étant pas assez longue pour couvrir leur parcours dans le secondaire.

Dans 86 % des cas où les répondants à l'enquête Génération sont retrouvés dans au moins deux fichiers annuels des apprenants, il n'y a pas de trous d'observation entre la première et la dernière année observée. Les trous de parcours sont plus nombreux pour les sortants de l'enseignement supérieur, ainsi que pour les composantes SIFA et Universités de la base de sondage. L'entrée en apprentissage dans une formation du supérieur peut être l'une des explications à ce constat quand cette entrée en apprentissage ne se fait pas

immédiatement après la sortie du secondaire. Dans ce cas, elle entraîne mécaniquement une sortie temporaire du champ des fichiers des apprenants. Par ailleurs, comme déjà indiqué, les passages par des établissements hors champ de SYSCA, notamment le privé hors contrat, les formations agricoles ou une scolarité à l'étranger, créent aussi des trous d'observation dans les fichiers des apprenants.

Tableau 7 • Nombre de présences dans les fichiers annuels des apprenants pour les répondants qui y sont retrouvés au moins une fois, selon le plus haut diplôme atteint

	Nombre de fichiers annuels dans lesquels les répondants retrouvés au moins une fois sont présents												Total
	1	2	3	4	5	6	7	8	9	10	11	12	
01-Non diplômés	3,7%	3,0%	2,4%	8,1%	21,1%	26,8%	21,1%	11,0%	2,5%	0,2%	0,0%	0,0%	100%
02-CAP-BEP-MC-Divers Niv.3	2,7%	2,9%	2,8%	2,6%	4,2%	30,8%	29,8%	16,4%	5,7%	1,4%	0,5%	0,1%	100%
03-Bac pro-BT-Divers Niv.4	2,5%	2,0%	1,4%	1,5%	2,3%	2,7%	36,8%	28,0%	14,9%	5,2%	2,0%	0,7%	100%
04-Bac technologique	1,6%	1,2%	0,5%	2,2%	1,8%	2,3%	37,8%	31,2%	15,1%	4,4%	1,6%	0,2%	100%
05-Bac général	2,4%	1,1%	1,3%	1,7%	2,6%	3,4%	52,1%	23,6%	8,3%	2,3%	0,9%	0,1%	100%
06-Bac+2 sauf santé/social	2,2%	3,4%	1,2%	1,2%	1,5%	3,2%	14,8%	7,6%	37,2%	19,6%	6,5%	1,6%	100%
07-Bac+2 à +4 santé/social	2,1%	2,1%	4,5%	4,1%	4,1%	8,3%	45,9%	11,6%	6,6%	4,5%	3,7%	2,5%	100%
08-Licence professionnelle	4,4%	1,5%	1,0%	2,8%	4,0%	4,4%	19,5%	17,3%	20,0%	16,7%	6,7%	1,6%	100%
09-Autre bac+3 et bac+4	1,7%	2,7%	3,4%	5,0%	8,3%	11,8%	45,6%	12,1%	5,1%	2,6%	1,2%	0,4%	100%
10-Bac+5 sauf écoles	3,6%	2,7%	4,8%	6,5%	13,6%	22,3%	28,1%	8,5%	6,5%	1,9%	1,0%	0,5%	100%
11-Ecoles (Ingé, commerce)	2,4%	2,8%	2,6%	2,6%	4,1%	12,1%	24,1%	16,0%	22,8%	5,7%	1,7%	3,0%	100%
12-Doctorat	28,7%	22,3%	15,5%	11,6%	10,8%	6,8%	4,4%	0,0%	0,0%	0,0%	0,0%	0,0%	100%

Champ : répondants (hors post-initiaux) retrouvés au moins une fois dans les fichiers des apprenants 2005 à 2016.

Note : les cases grisées correspondent aux valeurs supérieures à 10%. la pondération associée aux répondants de l'enquête Génération n'est pas utilisée ici.

Source : Céreq-Depp. Rapprochement de l'enquête Génération 2017 à trois ans et des fichiers des apprenants FAERE.

Tableau 8 • Première année observée pour les répondants à Génération 2017 retrouvés au moins une fois dans les fichiers des apprenants

1ère année observée pour les répondants à Génération 2017 retrouvés au moins une fois dans les fichiers des apprenants												
Selon la composante de la base de sondage	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
Sysca (second degré)	1%	2%	7%	13%	18%	30%	14%	5%	3%	2%	2%	2%
Sifa (apprentissage)	21%	7%	10%	16%	13%	10%	7%	2%	1%	2%	4%	9%
Universités	60%	11%	15%	7%	4%	1%	1%	1%	0%	0%	0%	0%
Autres établissements	79%	6%	6%	3%	3%	1%	1%	0%	0%	0%	0%	0%
DREES	41%	34%	18%	2%	1%	2%	2%	1%	0%	0%	0%	0%
Culture	95%	4%	0%	0%	2%	0%	0%	0%	0%	0%	0%	0%
DGAFP	93%	3%	1%	1%	1%	0%	1%	1%	0%	0%	0%	0%
Agriculture, Sport, Cifre, Contrats de prof.	s.o.	s.o.	s.o.	s.o.	s.o.	s.o.	s.o.	s.o.	s.o.	s.o.	s.o.	s.o.
Selon le plus haut diplôme atteint	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
01-Non diplômés	0%	0%	1%	3%	12%	21%	27%	20%	7%	2%	3%	4%
02-CAP-BEP-MC-Divers Niv.3	1%	1%	3%	7%	17%	30%	29%	2%	2%	2%	3%	3%
03-Bac pro-BT-Divers Niv.4	4%	3%	8%	15%	27%	34%	2%	1%	1%	1%	2%	3%
04-Bac technologique	6%	7%	13%	18%	28%	22%	3%	1%	0%	0%	1%	1%
05-Bac général	14%	11%	18%	23%	20%	8%	1%	1%	1%	1%	0%	2%
06-Bac+2 sauf santé/social	13%	14%	27%	36%	1%	1%	1%	1%	1%	1%	3%	2%
07-Bac+2 à +4 santé/social	47%	28%	16%	2%	0%	2%	1%	0%	0%	2%	0%	1%
08-Licence professionnelle	29%	26%	36%	1%	0%	1%	1%	1%	0%	1%	0%	4%
09-Autre bac+3 et bac+4	55%	20%	20%	1%	1%	1%	1%	1%	0%	0%	1%	0%
10-Bac+5 sauf écoles	93%	1%	1%	1%	1%	0%	0%	0%	0%	0%	0%	2%
11-Ecoles (Ingé, commerce)	90%	1%	1%	1%	1%	1%	1%	1%	0%	1%	1%	2%
12-Doctorat	99%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Ensemble	32%	6%	10%	10%	11%	14%	7%	3%	1%	1%	2%	2%

Champ : répondants (hors post-initiaux) retrouvés au moins une fois dans les fichiers des apprenants 2005 à 2016.

Source : Céreq-Depp. Rapprochement de l'enquête Génération 2017 à trois ans et des fichiers des apprenants FAERE.

Tableau 9 • Dernière année observée pour les répondants à Génération 2017 retrouvés au moins une fois dans les fichiers des apprenants

Dernière année observée pour les répondants à Génération 2017 retrouvés au moins une fois dans les fichiers des apprenants												
Selon la composante de la base de sondage	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
Sysca (second degré)	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	99%
Sifa (apprentissage)	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	100%
Universités	1%	2%	2%	3%	7%	12%	16%	9%	13%	9%	10%	14%
Autres établissements	2%	2%	2%	4%	7%	15%	26%	16%	18%	5%	4%	0%
DREES	0%	1%	1%	1%	5%	8%	13%	36%	32%	1%	1%	0%
Culture	7%	7%	4%	9%	24%	27%	18%	0%	2%	0%	2%	0%
DGAFP	2%	0%	2%	5%	8%	16%	17%	8%	38%	4%	2%	0%
Agriculture, Sport, Cifre, Contrats de prof.	s.o.	s.o.	s.o.	s.o.	s.o.	s.o.	s.o.	s.o.	s.o.	s.o.	s.o.	s.o.
Selon le plus haut diplôme atteint	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
01-Non diplômés	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	1%	98%
02-CAP-BEP-MC-Divers Niv.3	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	100%
03-Bac pro-BT-Divers Niv.4	0%	0%	0%	0%	0%	0%	0%	0%	1%	2%	2%	95%
04-Bac technologique	0%	0%	0%	0%	0%	0%	1%	3%	4%	11%	15%	65%
05-Bac général	0%	0%	0%	1%	2%	2%	4%	8%	14%	20%	24%	25%
06-Bac+2 sauf santé/social	0%	0%	0%	0%	0%	1%	2%	4%	6%	9%	4%	73%
07-Bac+2 à +4 santé/social	0%	0%	0%	2%	4%	7%	17%	26%	21%	1%	0%	20%
08-Licence professionnelle	0%	0%	0%	1%	2%	4%	3%	7%	14%	5%	16%	47%
09-Autre bac+3 et bac+4	1%	1%	2%	4%	7%	11%	16%	22%	22%	3%	4%	7%
10-Bac+5 sauf écoles	1%	2%	3%	6%	13%	21%	25%	5%	5%	1%	1%	19%
11-Ecoles (Ingé, commerce)	0%	1%	1%	2%	3%	10%	22%	14%	21%	2%	0%	24%
12-Doctorat	28%	22%	15%	12%	10%	7%	5%	0%	0%	0%	1%	0%
Ensemble	1%	1%	1%	2%	3%	6%	8%	5%	6%	3%	4%	61%

Champ : répondants (hors post-initiaux) retrouvés au moins une fois dans les fichiers des apprenants 2005 à 2016.

Source : Céreq-Depp. Rapprochement de l'enquête Génération 2017 à trois ans et des fichiers des apprenants FAERE.

Tableau 10 • Nombre d'années « vides » entre la première et la dernière année observée pour les répondants à Génération 2017 retrouvés au moins deux fois dans les fichiers des apprenants

	Nombre d'années "manquantes" entre la première année et la dernière année d'observation									
Selon la composante de la base de sondage	0	1	2	3	4	5	6	7	8	9
Sysca (second degré)	94%	3%	1%	1%	0%	0%	0%	0%	0%	0%
Sifa (apprentissage)	62%	12%	10%	7%	4%	3%	2%	1%	0%	0%
Universités	86%	3%	3%	2%	2%	2%	1%	0%	0%	0%
Autres établissements	96%	2%	1%	0%	1%	0%	0%	0%	0%	0%
DREES	97%	2%	1%	0%	0%	0%	0%	0%	0%	0%
Culture	96%	4%	0%	0%	0%	0%	0%	0%	0%	0%
DGAFF	94%	4%	1%	1%	0%	1%	0%	0%	0%	0%
Agriculture, Sport, Cifre, Contrats de prof.	s.o.	s.o.	s.o.	s.o.	s.o.	s.o.	s.o.	s.o.	s.o.	s.o.
Selon le plus haut diplôme atteint	0	1	2	3	4	5	6	7	8	9
01-Non diplômés	94%	3%	2%	1%	0%	0%	0%	0%	0%	0%
02-CAP-BEP-MC-Divers Niv.3	88%	5%	4%	1%	1%	0%	0%	0%	0%	0%
03-Bac pro-BT-Divers Niv.4	90%	5%	2%	1%	1%	0%	0%	0%	0%	0%
04-Bac technologique	92%	5%	1%	1%	1%	0%	0%	0%	0%	0%
05-Bac général	94%	3%	1%	1%	0%	0%	0%	0%	0%	0%
06-Bac+2 sauf santé/social	82%	10%	4%	3%	1%	0%	0%	0%	0%	0%
07-Bac+2 à +4 santé/social	89%	3%	2%	1%	1%	1%	1%	0%	0%	0%
08-Licence professionnelle	68%	7%	12%	7%	3%	2%	1%	1%	0%	0%
09-Autre bac+3 et bac+4	94%	2%	1%	1%	0%	1%	1%	0%	0%	0%
10-Bac+5 sauf écoles	82%	2%	2%	3%	5%	4%	2%	1%	0%	0%
11-Ecoles (Ingé, commerce)	80%	4%	6%	5%	2%	2%	1%	0%	0%	0%
12-Doctorat	98%	1%	0%	0%	0%	0%	1%	0%	0%	0%
Ensemble	86%	5%	3%	2%	1%	1%	1%	0%	0%	0%

Champ : répondants (hors post-initiaux) retrouvés au moins deux fois dans les fichiers des apprenants 2005 à 2016.

Source : Céreq-Depp. Rapprochement de l'enquête Génération 2017 à trois ans et des fichiers des apprenants FAERE.

4. Reconstruire les étapes principales des parcours d'inscription avec les données administratives

Comme rappelé dans la partie 1.1.b, les fichiers des apprenants permettent de connaître, dans le champ des établissements couverts par SYSCA et SIFA :

- la classe d'inscription (nature, niveau, spécialité),
- le diplôme préparé,
- le secteur de l'établissement d'inscription (public ou privé²⁹),
- le statut de l'apprenant (essentiellement pour distinguer les apprentis des scolaires, distinction qui ne coïncide pas avec le partage entre SYSCA et SIFA),
- la date de sortie et le motif de sortie en cas de départ en cours d'année scolaire (pour la partie SYSCA depuis 2011 uniquement),
- les inscriptions et résultats à certains examens.

Par ailleurs, une variable permet de repérer les boursiers, mais, selon la documentation de la DEPP³⁰, l'information ne serait pas fiable en raison du calendrier de remontée de l'information et cette information n'est pas présente dans SIFA. Selon la documentation transmise, une variable permet également d'identifier le régime d'inscription de l'élève par rapport à la demi-pension et l'hébergement (internat), mais cette information n'a pas été demandée, ni transmise.

Parmi ces différentes variables, certaines sont intéressantes à exploiter pour caractériser les parcours, mais elles nécessitent la construction de variables synthétiques pour en faciliter l'usage à des fins d'analyse. C'est en particulier le cas de la classe d'inscription. Cette construction impose des choix subjectifs, qui peuvent évoluer dans le temps, notamment du fait de l'évolution des politiques publiques. De ce fait, en cas de renouvellement de l'exercice, il serait nécessaire de suivre précisément l'évolution des nomenclatures de référence et des effectifs associés.

4.1. Construire une variable synthétique sur la classe d'inscription

Les fichiers des apprenants proposent deux variables très détaillées pour définir la classe d'inscription :

- La variable **MEFST11** : elle s'appuie sur une nomenclature propre à SYSCA, appelée MEFSTAT (Module Élémentaire de Formation STATistique), qui vise à décrire l'ensemble des formations présentes dans les établissements du second degré. Cette nomenclature est hiérarchisée (voir schéma *infra*). La variable est à blanc pour les élèves issus de SIFA.

Schéma de construction du code « MEFSTAT »

CARACTERES	NIVEAU D'INFORMATION
① 1er caractère	“ Degré d'enseignement ” : 1er degré, 2nd degré, post-Bac,
② 2ème caractère	“ Cycle d'enseignement ” : 1er cycle, 2nd cycle G & T, CPGE, STS..
③ 3ème caractère	“ Dispositif de formation ” : CAP 1 an, CAP 2ans, ..., BTS 2 ans, ...
④ 4ème caractère	“ Classe ” : 6e, 5e, ..., 2nde Pro, ..., 1ère année CPGE...
⑤ 5ème caractère	“ Filière ” du 2nd cycle G&T : Voie Générale / Voie Techno / BT
⑥ 6ème caractère	- “ Série ” du bac G & T : S, L, ES, STI, STG, - “ Type de CPGE ” : Scientifique, Littéraire, Eco & Commerciales

Source : DEPP-Documentation des fichiers FAERE des apprenants.

²⁹ Une variable permet de distinguer les inscriptions dans des établissements privés sous contrat des inscriptions dans des établissements hors contrat, mais ces derniers ne sont pas couverts de façon exhaustive. La documentation de la DEPP précise qu'un établissement du second degré est considéré « sous contrat » dès lors qu'il comporte au moins une division sous contrat. L'information retenue dans le fichier FAERE porte sur l'existence d'un contrat avec l'établissement, que l'élève considéré soit inscrit ou non dans la division couverte par ce contrat.

³⁰ Dans l'extrait du fichier des apprenants 2016 transmis au Céreq, l'information est renseignée pour 62 % des individus et, parmi eux, 30 % sont identifiés comme boursiers.

- La variable **FORTRAMS_7** : elle s'appuie sur une nomenclature transversale aux différents systèmes d'information et en particulier, à SYSCA et SIFA. Selon la documentation de la DEPP, « la FORMation TRAnsversale de la Maternelle au Supérieur » a pour objet de décrire, à travers une grille de lecture unique, l'ensemble des formations de la petite section de maternelle au doctorat. Cette nomenclature s'appuie sur des critères permettant de caractériser au mieux l'ensemble des formations dispensées au sein du système éducatif et répondant aux besoins des différents partenaires. Pour construire cette nomenclature, les dimensions d'analyse présentes dans la Classification Internationale Type de l'Éducation (CITE), rénovée en 1997, ont également été prises en compte ». Cette nomenclature est également hiérarchisée (voir schéma *infra*).

Schéma de construction du code « FORTRAMS »

CARACTERES		NIVEAU D'INFORMATION
①	1er caractère	« Niveau d'éducation »
②	2ème caractère	« Dominante de la formation »
③	3ème caractère	« Finalité de la formation »
④	4ème caractère	Premier « N° d'ordre » (non significatif en tant que tel)
⑤	5ème caractère	Deuxième « N° d'ordre » (non significatif en tant que tel)
⑥	6ème caractère	« Durée de la formation »
⑦	7ème caractère	« Année dans la formation »

FORTRAMS_7 :	« Classe », « Diplôme », « Prépa », ... + Année dans la formation
FORTRAMS_6 :	« Classe », « Diplôme », « Prépa », ... + Durée de la formation
FORTRAMS_5 :	« Classe » (6e-3e, ...), « Diplôme » (BEP, DUT, ...), « Prépa » (CPGE, ...), ...
FORTRAMS_4 :	« Classe » (6e-3e, ...), « Diplôme » (BEP, DUT, ..), « Prépa », Form Géné/Tec/Pro/T-Pro
FORTRAMS_3 :	+ « Diplômante ou non »
FORTRAMS_2 :	Formation « GENE/TEC/PRO »
FORTRAMS_1 :	Niveau d'Educ.

Source : DEPP-Documentation des fichiers FAERE des apprenants.

Nous avons fait le choix de nous appuyer sur la nomenclature FORMTRAMS puisqu'elle est commune aux deux sources qui alimentent le fichier des apprenants. La nomenclature détaillée est présentée dans l'annexe 1.A (hors 1er degré). Les annexes 1.B et 1.C présentent les regroupements que nous avons opérés pour étudier les parcours. Parmi les points à noter à leurs sujets :

- au niveau du collège, nous avons distingué les quatre classes, de la sixième à la troisième, sans y intégrer les classes de SEGPA et nous avons identifié le passage par une SEGPA au collège de façon globale, sans différencier les niveaux de classes. Un autre choix possible aurait été de regrouper les classes de SEGPA avec les classes équivalentes et de créer une variable indicatrice de présence en SEGPA (une « inscription au moins un an », par exemple).
- au niveau du lycée, nous avons distingué la voie générale et la voie technologique pour les premières et les terminales, sans aller plus dans le détail pour la spécialité même si la nomenclature le permet.
- pour les CAP, la nomenclature propose un détail qui prend en compte la durée du cycle, soit six modalités : 1^{ère} année sur un cycle d'un an ; 1^{ère} année sur un cycle de 2 ans ; 2^{ème} année sur un cycle de 2 ans ; 1^{ère} année sur un cycle de 3 ans ; etc. Nous avons fait le choix de nous limiter à deux modalités : classe non terminale ou classe terminale du cycle. Là aussi, aucun détail n'a été fait pour les spécialités.
- pour les baccalauréats professionnels, les BEP, les brevets professionnels et les BTS, nous avons procédé comme pour le CAP.

Tableau 11 • Répartition des effectifs des fichiers des apprenants 2016 et 2005 selon la variable de classe synthétique que nous avons construite

	Dans le fichier 2016		Dans le fichier 2005	
	Effectif	Répartition (en %)	Effectif	Répartition (en %)
02-6 ^e	2	0,0%	10 373	32,3%
03-5e	9	0,0%	8 818	27,5%
04-4e générale	3 622	2,1%	5 656	17,6%
05-3e générale	11 779	6,8%	3 306	10,3%
06-SES/SEGPA	1 481	0,9%	180	0,6%
07-Autres niveau collège	22	0,0%	42	0,1%
08-2de générale	2 567	1,5%	1 672	5,2%
09-1re générale	1 229	0,7%	840	2,6%
10-1re technologique	1 019	0,6%	118	0,4%
11-Terminale générale	5 480	3,2%	659	2,1%
12-Terminale technologique	6 238	3,6%	77	0,2%
13-Autres niveau lycée	1 323	0,8%	9	0,0%
14-CAP, classe non terminale	15 583	9,0%	16	0,0%
15-CAP, classe terminale	35 519	20,5%	4	0,0%
20-BEP, classe non terminale		0,0%	109	0,3%
21-BEP, classe terminale	13 402	7,7%	54	0,2%
23-Bac pro, classe non terminale		0,0%	12	0,0%
24-Bac pro, classe terminale	35 156	20,3%	13	0,0%
29-Mention complémentaire	1 305	0,8%		0,0%
30-Brevet pro, classe non terminale	1 724	1,0%		0,0%
31-Brevet pro, classe terminale	5 218	3,0%		0,0%
35-Fortrams1=4	341	0,2%	1	0,0%
36-CPGE 1/2	116	0,1%	58	0,2%
37-CPGE 2/2	869	0,5%	43	0,1%
38-BTS, classe non terminale	5 380	3,1%	20	0,1%
39-BTS, classe terminale	14 461	8,4%	16	0,0%
44-DUT 1/2	64	0,0%		0,0%
45-DUT 2/2	325	0,2%		0,0%
46-Autre Fortrams1=5	475	0,3%	1	0,0%
47-Licence pro	2 189	1,3%		0,0%
48-Master pro 1/2	170	0,1%		0,0%
49-Master pro 2/2	1 931	1,1%		0,0%
53-Ingenieur DIV	1 497	0,9%		0,0%
54-Commerce	4	0,0%		0,0%
56-Commerce TERM	16	0,0%		0,0%
57-Commerce DIV	1 318	0,8%		0,0%
58-Autre Fortrams1=6	1 181	0,7%		0,0%
99-Autres SIFA	10	0,0%		0,0%
Total	173 025	100,0%	32 097	100,0%

Note : la répartition couvre l'ensemble de l'échantillon de l'enquête Génération 2017, répondants et non répondants.
Sources : DEPP-Extraits des fichiers des apprenants 2005 et 2016.

La variable de classe construite à partir de la variable FORTRANS_7 comprend 44 modalités, dont certaines ne sont pas présentes dans les fichiers des apprenants, mais figurent dans la nomenclature FORMTRANS_7 (classes préélémentaires et élémentaires). Le tableau 11 présente la répartition des effectifs des extraits des fichiers des apprenants transmis au Céreq selon cette variable synthétique pour deux années : le fichier le plus récent de la période étudiée (2016) et le fichier le plus ancien (2005). Cette répartition porte sur l'intégralité de l'échantillon de l'enquête Génération, répondants et non-répondants. De façon attendue en raison du champ de l'enquête Génération, les niveaux terminaux sont les plus représentés dans le fichier 2016, année théorique³¹ de sortie du système éducatif pour la construction de la base de sondage (71 % de l'échantillon). Dans le fichier 2005, le plus ancien mobilisé, l'essentiel des effectifs retrouvés est inscrit au collège (88 % de l'échantillon). Cette année-là, près de 22 000 personnes de l'échantillon, sont retrouvés dans une classe postérieure à la sixième. Pour ces personnes, une profondeur historique supérieure à celle retenue serait donc nécessaire pour remonter à la classe de sixième.

4.2. Repérer les classes d'inscription successives des répondants

Les tableaux 12.a et 12.b présentent la proportion de répondants à l'enquête Génération pour lesquels au moins une inscription dans une classe donnée est disponible dans les fichiers administratifs mobilisés ; d'une part, selon la composante de la base de sondage ; d'autre part, selon le plus haut diplôme atteint. Ces classes sont définies selon les modalités de la variable synthétique présentée dans la partie 4.1. En particulier, pour les classes de collège, les SEGPA ne sont pas prises en compte. Pour le lycée, les premières et terminales regroupent les filières générales et technologiques.

Les résultats présentés ne sont calculés que sur un champ restreint aux seuls répondants retrouvés au moins une fois dans les fichiers de la DEPP. Pour obtenir la proportion totale de répondants de l'enquête Génération pour lesquels l'information considérée est disponible dans les données administratives mobilisées, il faut donc multiplier les proportions indiquées par la proportion de répondants repérés au moins une fois qui figure dans la colonne grisée à droite des tableaux. Par exemple, pour les répondants du segment « Universités » de la base de sondage, la classe de terminale est retrouvée dans les fichiers de la DEPP pour 88 %*92 %=81 % des répondants relevant de ce segment.

Tableau 12.a • Repérage des inscriptions en classe de collège et de lycée, selon la composante de la base de sondage

	Proportion de répondants à Génération 2017 pour lesquels la classe considérée est retrouvée dans les fichiers des apprenants 2005-2016 (parmi les répondants retrouvés au moins une fois dans les fichiers de la DEPP)							% de repérés au moins une fois dans les fichiers de la DEPP
	6e	5e	4e	3e	2de	1ère	Term.	
Sysca (second degré)	82%	81%	82%	83%	24%	21%	19%	100%
Sifa (apprentissage)	66%	72%	74%	73%	30%	27%	26%	100%
Universités	54%	71%	83%	89%	88%	91%	92%	88%
Autres établissements	41%	66%	80%	88%	89%	93%	94%	56%
DREES	73%	84%	90%	91%	88%	95%	97%	20%
Culture	18%	35%	55%	73%	80%	85%	93%	37%
DGAFP	35%	61%	77%	89%	92%	94%	96%	41%
Agriculture, Sport, Cifre, Contrats de prof.	s.o.	s.o.	s.o.	s.o.	s.o.	s.o.	s.o.	0%
Ensemble	66%	75%	81%	84%	53%	52%	52%	74%

Champ : répondants (hors post-initiaux).

Note : la pondération associée aux répondants de l'enquête Génération n'est pas utilisée ici.

Source : Céreq-Depp. Rapprochement de l'enquête Génération 2017 à trois ans et des fichiers des apprenants FAERE.

³¹ On rappelle que l'appartenance effective au champ de l'enquête n'est déterminée que pour les répondants, grâce à la passation d'un questionnaire filtre lors de l'enquête.

Tableau 12.b • Repérage des inscriptions en classe de collège et de lycée, selon le plus haut diplôme atteint

	Proportion de répondants à Génération 2017 pour lesquels la classe considérée est retrouvée dans les fichiers des apprenants 2005-2016, (parmi les répondants retrouvés au moins une fois dans les fichiers de la DEPP)							% de repérés au moins une fois dans les fichiers de la DEPP
	6e	5e	4e	3e	2de	1ère	Term.	
01-Non diplômés	78%	77%	77%	77%	19%	13%	8%	96%
02-CAP-BEP-MC-Divers Niv.3	68%	67%	66%	64%	4%	1%	1%	94%
03-Bac pro-BT-Divers Niv.4	86%	87%	88%	88%	10%	4%	3%	75%
04-Bac technologique	88%	90%	92%	94%	92%	97%	97%	86%
05-Bac général	83%	87%	91%	93%	95%	96%	97%	85%
06-Bac+2 sauf santé/social	83%	86%	89%	91%	65%	65%	64%	73%
07-Bac+2 à +4 santé/social	68%	78%	85%	88%	86%	91%	93%	39%
08-Licence professionnelle	75%	83%	87%	91%	81%	81%	81%	68%
09-Autre bac+3 et bac+4	55%	70%	81%	89%	91%	93%	95%	65%
10-Bac+5 sauf écoles	29%	55%	76%	85%	90%	93%	95%	68%
11-Ecoles (Ingé., commerce)	37%	70%	84%	89%	90%	92%	92%	66%
12-Doctorat	0%	2%	8%	22%	39%	65%	92%	48%
Ensemble	66%	75%	81%	84%	53%	52%	52%	74%

Champ : répondants (hors post-initiaux).

Note : la pondération associée aux répondants de l'enquête Génération n'est pas utilisée ici.

Source : Céreq-Depp. Rapprochement de l'enquête Génération 2017 à trois ans et des fichiers des apprenants FAERE.

Ces résultats montrent que pour les répondants retrouvés au moins une fois dans les fichiers des apprenants, il est généralement possible de retrouver les classes successives suivies au collège et, quand il y a lieu, au lycée. Cependant, une profondeur historique plus importante serait nécessaire pour remonter à la sixième pour l'ensemble des sortants du supérieur, en laissant de côté les doctorats, qui nécessiteraient une profondeur encore plus grande.

Les tableaux 14.a et 14.b, qui présentent la première classe connue à partir des fichiers de la DEPP, confortent cette dernière conclusion de façon plus directe puisque, pour les répondants retrouvés au moins une fois dans les données administratives, la première classe repérée est la sixième dans 83 % des cas pour les jeunes ayant pour plus haut diplôme un bac général, 86 % pour les bacs technologiques et 88 % pour les baccalauréats professionnels et autres diplômes de niveau 4. Ce taux est de 75 % au niveau des licences professionnelles, 37 % pour les diplômés d'écoles d'ingénieurs ou de commerce et 29 % pour les autres bacs+5. En allongeant de trois à quatre ans la profondeur historique des fichiers administratifs mobilisés, on pourrait atteindre une connaissance de la sixième suivie (et des éventuels redoublements au collège) pour neuf répondants sur dix³², quel que soit leur plus haut diplôme, excepté les titulaires d'un doctorat, pour lesquels la classe de sixième ne serait récupérée que dans un cas sur cinq³³.

Pour les jeunes non diplômés ou ayant comme plus haut diplôme un CAP ou BEP, la moindre connaissance apparente des classes de la sixième à la troisième relativement aux bacheliers, peut s'expliquer par la façon de construire la variable synthétique de classes. Celle-ci isole et regroupe ensemble les classes de SEGPA. Or, 9 % des non-diplômés repérés dans les fichiers de la DEPP ont pour première classe repérée une classe de SEGPA. Pour ceux ayant comme plus haut diplôme un CAP ou BEP, c'est 16 %.

L'examen de ces tableaux, ainsi que des tableaux 13a et 13 b, portant sur le repérage d'inscription dans divers types de classes (CAP, BEP, bac pro...), et les tableaux 15a et 15b, qui portent sur la dernière classe retrouvée dans les fichiers administratifs mobilisés, vont dans le même sens. Ils montrent également que la nature des classes enregistrées dans les fichiers administratifs est cohérente avec les plus hauts diplômes atteints. Par exemple, il y a très peu d'inscriptions en seconde repérées pour les jeunes ayant pour plus haut diplôme un

³² Sous réserve de les retrouver au moins une fois dans les fichiers de la DEPP.³³ *Idem*.

CAP ou BEP (seulement 4 % des cas). Une inscription en terminale générale ou technologique est repérée dans plus de neuf cas sur dix pour les sortants du supérieur et pour les bacheliers généraux et technologiques, avec deux bémols : les bacs+2 hors santé/social (64 % semblent passés par une terminale générale ou technologique) et les licences professionnelles (81 %), mais ces niveaux accueillent beaucoup de jeunes passés par la filière professionnelle. Etc.

Ces résultats tendent à confirmer une bonne capacité à reconstituer des parcours avec les données administratives, sur les champs qu'elles couvrent, dès lors que les personnes sont retrouvées au moins une fois donc, en particulier, dès lors qu'un INE est disponible lors de la construction de la base de sondage.

Tableau 13.a • Poursuite du parcours scolaire, selon la composante de la base de sondage

	Parmi les répondants à Génération 2017 retrouvés au moins une fois dans les fichiers de la DEPP, proportion de ceux ayant une inscription en ...					% de repérés au moins une fois dans les fichiers de la DEPP
	CAP	BEP	Bac pro	Classe préparatoire	BTS	
Sysca (second degré)	26%	0%	57%	0%	21%	100%
Sifa (apprentissage)	48%	2%	34%	3%	23%	100%
Universités	1%	1%	5%	10%	14%	88%
Autres établissements	1%	1%	3%	30%	6%	56%
DREES	0%	4%	1%	1%	3%	20%
Culture	0%	2%	4%	2%	11%	37%
DGAFF	0%	0%	0%	64%	2%	41%
Agriculture, Sport, Cifre, Contrats de prof.	s.o.	s.o.	s.o.	7%	s.o.	0%
Ensemble	18%	1%	30%	6%	17%	74%

Champ : répondants (hors post-initiaux).

Note : la pondération associée aux répondants de l'enquête Génération n'est pas utilisée ici.

Source : Céreq-Depp. Rapprochement de l'enquête Génération 2017 à trois ans et des fichiers des apprenants FAERE.

Tableau 13.b • Poursuite du parcours scolaire, selon le plus haut diplôme atteint

	Parmi les répondants à Génération 2017 retrouvés au moins une fois dans les fichiers de la DEPP, proportion de ceux ayant une inscription en ...					% de repérés au moins une fois dans les fichiers de la DEPP
	CAP	BEP	Bac pro	Classe préparatoire	BTS	
01-Non diplômés	37%	0%	42%	0%	0%	96%
02-CAP-BEP-MC-Divers Niv.3	78%	0%	36%	0%	0%	35%
03-Bac pro-BT-Divers Niv.4	20%	1%	89%	0%	13%	75%
04-Bac technologique	4%	1%	3%	1%	22%	86%
05-Bac général	5%	0%	1%	4%	7%	85%
06-Bac+2 sauf santé/social	3%	1%	28%	1%	74%	73%
07-Bac+2 à +4 santé/social	0%	5%	2%	1%	14%	39%
08-Licence professionnelle	1%	1%	9%	2%	44%	68%
09-Autre bac+3 et bac+4	1%	1%	3%	6%	11%	65%
10-Bac+5 sauf écoles	0%	1%	1%	12%	8%	68%
11-Ecoles (Ingé, commerce)	0%	1%	1%	44%	7%	66%
12-Doctorat	0%	0%	0%	26%	1%	48%
Ensemble	18%	1%	30%	6%	17%	74%

Champ : répondants (hors post-initiaux).

Note : la pondération associée aux répondants de l'enquête Génération n'est pas utilisée ici.

Source : Céreq-Depp. Rapprochement de l'enquête Génération 2017 à trois ans et des fichiers des apprenants FAERE.

Tableau 14.a • Première classe retrouvée dans les fichiers de la DEPP, selon la composante de la base de sondage

	Classe d'inscription lors de la première année de présence dans les fichiers FAERE 2005 à 2016 (pour les répondants à Génération 2017 retrouvés au moins une fois dans ces fichiers)											% de repérés au moins une fois dans les fichiers de la DEPP
	6e	5e	4e	3e	SES/SEGPA	Autres niv. Collège	2de	1ère	Terminale	Autres niv. Lycée	Autres (>)	
Sysca (second degré)	81%	1%	1%	1%	7%	1%	1%	0%	0%	1%	5%	100%
Sifa (apprentissage)	66%	7%	4%	2%	3%	1%	1%	0%	0%	1%	15%	100%
Universités	54%	17%	12%	6%	0%	0%	4%	2%	2%	0%	2%	88%
Autres établissements	41%	25%	14%	8%	0%	0%	5%	3%	2%	0%	2%	56%
DREES	73%	11%	6%	1%	0%	0%	4%	1%	1%	0%	3%	20%
Culture	18%	16%	20%	18%	0%	0%	11%	5%	9%	2%	0%	37%
DGAFP	35%	26%	17%	12%	0%	0%	4%	3%	3%	0%	2%	41%
Agriculture, Sport, Cifre, Contrats de prof.	s.o.	s.o.	s.o.	s.o.	s.o.	s.o.	s.o.	s.o.	s.o.	s.o.	s.o.	0%
Ensemble	66%	10%	7%	4%	3%	1%	2%	1%	1%	0%	5%	74%

Champ : répondants (hors post-initiaux).

Note : la pondération associée aux répondants de l'enquête Génération n'est pas utilisée ici.

Source : Céreq-Depp. Rapprochement de l'enquête Génération 2017 à trois ans et des fichiers des apprenants FAERE.

Tableau 14.b • Première classe retrouvée dans les fichiers de la DEPP, selon le plus diplôme atteint

	Classe d'inscription lors de la première année de présence dans les fichiers FAERE 2005 à 2016 (pour les répondants à Génération 2017 retrouvés au moins une fois dans ces fichiers)											% de repérés au moins une fois dans les fichiers de la DEPP
	6e	5e	4e	3e	SES/SEGPA	Autres niv. Collège	2de	1ère	Terminale	Autres niv. Lycée	Autres (>)	
01-Non diplômés	78%	1%	1%	2%	9%	2%	1%	1%	0%	1%	4%	96%
02-CAP-BEP-MC-Divers Niv.3	68%	1%	2%	2%	16%	3%	0%	0%	0%	1%	7%	35%
03-Bac pro-BT-Divers Niv.4	86%	2%	1%	1%	2%	0%	0%	0%	0%	1%	6%	75%
04-Bac technologique	88%	3%	2%	2%	0%	0%	2%	0%	0%	1%	3%	86%
05-Bac général	83%	5%	3%	2%	0%	0%	2%	1%	1%	0%	2%	85%
06-Bac+2 sauf santé/social	83%	3%	3%	2%	0%	0%	1%	0%	0%	0%	7%	73%
07-Bac+2 à +4 santé/social	68%	10%	7%	3%	0%	0%	5%	0%	1%	0%	6%	39%
08-Licence professionnelle	75%	8%	4%	4%	0%	0%	1%	1%	0%	0%	7%	68%
09-Autre bac+3 et bac+4	55%	15%	11%	8%	0%	0%	5%	2%	2%	0%	2%	65%
10-Bac+5 sauf écoles	29%	27%	20%	10%	0%	0%	6%	3%	1%	0%	4%	68%
11-Ecoles (Ingé, commerce)	37%	33%	14%	5%	0%	0%	3%	1%	0%	0%	6%	66%
12-Doctorat	0%	2%	6%	13%	0%	0%	19%	26%	27%	0%	6%	48%
Génération 2017	66%	10%	7%	4%	3%	1%	2%	1%	1%	0%	5%	74%

Champ : répondants (hors post-initiaux).

Note : la pondération associée aux répondants de l'enquête Génération n'est pas utilisée ici.

Source : Céreq-Depp. Rapprochement de l'enquête Génération 2017 à trois ans et des fichiers des apprenants FAERE.

Tableau 15.a • Dernière classe retrouvée dans les fichiers de la DEPP, selon la composante de la base de sondage

	Classe d'inscription lors de la dernière année de présence dans les fichiers FAERE 2005 à 2016 (pour les répondants à Génération 2017 retrouvés au moins une fois dans ces fichiers)											% de repérés au moins une fois dans les fichiers de la DEPP
	Classe de 6e à 1ère (ie SES/SEGPA)	Terminale	CAP (année terminale ou non)	Bac pro (année terminale ou non)	Brevet pro (année terminale ou non)	CGPE	BTS (année terminale ou non)	DUT (année terminale ou non)	Licence pro	Master pro (année terminale ou non)	Autres	
Sysca (second degré)	4%	7%	20%	45%	0%	0%	21%	0%	0%	0%	3%	100%
Sifa (apprentissage)	0%	0%	33%	14%	12%	0%	16%	0%	3%	4%	18%	100%
Universités	1%	63%	0%	3%	0%	2%	9%	1%	6%	6%	9%	88%
Autres établissements	2%	58%	0%	2%	0%	4%	5%	0%	0%	0%	29%	56%
DREES	3%	78%	0%	0%	1%	0%	3%	0%	1%	0%	14%	20%
Culture	2%	76%	0%	4%	0%	0%	11%	0%	0%	0%	7%	37%
DGAFP	0%	32%	0%	0%	0%	3%	1%	0%	0%	1%	63%	41%
Agriculture, Sport, Cifre, Contrats de prof.	s.o.	s.o.	s.o.	s.o.	s.o.	s.o.	s.o.	s.o.	s.o.	s.o.	s.o.	0%
Ensemble	2%	30%	13%	21%	2%	1%	15%	0%	3%	3%	10%	74%

Champ : répondants (hors post-initiaux).

Note : la pondération associée aux répondants de l'enquête Génération n'est pas utilisée ici.

Source : Céreq-Depp. Rapprochement de l'enquête Génération 2017 à trois ans et des fichiers des apprenants FAERE.

Tableau 15.b • Dernière classe retrouvée dans les fichiers de la DEPP, selon le plus diplôme atteint

	Classe d'inscription lors de la dernière année de présence dans les fichiers FAERE 2005 à 2016 (pour les répondants à Génération 2017 retrouvés au moins une fois dans ces fichiers)											% de repérés au moins une fois dans les fichiers de la DEPP
	Classe de 6 ^e à 1 ^{ère} (ie SES/SEG PA)	Termi-nale	CAP (année terminale ou non)	Bac pro (année terminale ou non)	Brevet pro (année terminale ou non)	CGPE	BTS (année terminale ou non)	DUT (année terminale ou non)	Licence pro	Master pro (année terminale ou non)	Autres	
01-Non diplômés	18%	8%	35%	39%	0%	0%	0%	0%	0%	0%	0%	96%
02-CAP-BEP-MC-Divers Niv.3	0%	0%	66%	29%	3%	0%	0%	0%	0%	0%	2%	35%
03-Bac pro-BT-Divers Niv.4	1%	0%	6%	70%	8%	0%	12%	0%	0%	0%	4%	75%
04-Bac technologique	0%	69%	3%	1%	1%	1%	21%	0%	0%	0%	4%	86%
05-Bac général	0%	82%	4%	0%	1%	1%	6%	0%	0%	0%	4%	85%
06-Bac+2 sauf santé/social	0%	19%	1%	1%	0%	1%	71%	2%	1%	0%	4%	73%
07-Bac+2 à +4 santé/social	1%	66%	0%	1%	0%	0%	4%	0%	0%	0%	27%	39%
08-Licence professionnelle	1%	27%	0%	1%	0%	1%	22%	0%	44%	0%	3%	68%
09-Autre bac+3 et bac+4	1%	76%	1%	1%	0%	2%	9%	0%	0%	0%	8%	65%
10-Bac+5 sauf écoles	1%	62%	0%	0%	0%	3%	5%	0%	0%	16%	11%	68%
11-Ecoles (Ingé, commerce)	1%	33%	0%	0%	0%	3%	2%	0%	0%	0%	61%	66%
12-Doctorat	1%	70%	0%	0%	0%	4%	1%	0%	0%	1%	23%	48%
Ensemble	2%	30%	13%	21%	2%	1%	15%	0%	3%	3%	10%	74%

Champ : répondants (hors post-initiaux).

Note : la pondération associée aux répondants de l'enquête Génération n'est pas utilisée ici.

Source : Céreq-Depp. Rapprochement de l'enquête Génération 2017 à trois ans et des fichiers des apprenants FAERE.

4.3 Repérer les redoublements dans les données administratives et comparer avec ce que dit l'enquête

Le repérage des redoublements au collège dans les fichiers des apprenants mobilisés s'effectue, à partir du nombre de fois où l'individu est inscrit respectivement en sixième, cinquième, quatrième et troisième, à partir de la variable synthétique de classe présentée dans la partie 4.1. Pour les redoublements au lycée, il en est de même, avec les classes de seconde, première et terminale. Le nombre de redoublements ainsi repéré à partir des données administratives est nécessairement inférieur à la réalité, puisqu'elles couvrent de manière incomplète le champ de l'enquête Génération et parce que la profondeur historique retenue est insuffisante pour couvrir l'intégralité de la scolarité dans le secondaire pour les jeunes sortis avec les plus hauts niveaux de formation.

Tableau 16.a • Redoublements repérés dans les fichiers de la DEPP, selon la composante de la base de sondage

	Proportion de répondants à Génération 2017 concernés parmi ceux retrouvés au moins une fois dans les fichiers de la DEPP			% de repérés au moins une fois dans les fichiers de la DEPP
	Redoublement au collège	Redoublement au Lycée	A été en classe de collège SEGPA ou assimilé	
Sysca (second degré)	15%	7%	9%	100%
Sifa (apprentissage)	13%	6%	4%	100%
Universités	4%	15%	0%	88%
Autres établissements	2%	12%	0%	56%
DREES	1%	14%	0%	20%
Culture	2%	5%	0%	37%
DGAFF	2%	4%	0%	41%
Agriculture, Sport, Cifre, Contrats de prof.	s.o.	s.o.	s.o.	0%
Ensemble	10%	10%	4%	74%

Champ : répondants (hors post-initiaux).

Note : la pondération associée aux répondants de l'enquête Génération n'est pas utilisée ici.

Source : Céreq-Depp. Rapprochement de l'enquête Génération 2017 à trois ans et des fichiers des apprenants FAERE.

Tableau 16.b • Redoublements repérés dans les fichiers de la DEPP, selon la composante de la base de sondage

	Proportion de répondants à Génération 2017 concernés parmi ceux retrouvés au moins une fois dans les fichiers de la DEPP			% de repérés au moins une fois dans les fichiers de la DEPP
	Redoublement au collège	Redoublement au Lycée	A été en classe de collège SEGPA ou assimilé	
01-Non diplômés	19%	9%	13%	96%
02-CAP-BEP-MC-Divers Niv.3	16%	1%	20%	94%
03-Bac pro-BT-Divers Niv.4	17%	2%	3%	75%
04-Bac technologique	12%	30%	0%	86%
05-Bac général	5%	29%	0%	85%
06-Bac+2 sauf santé/social	9%	16%	0%	73%
07-Bac+2 à +4 santé/social	2%	16%	0%	39%
08-Licence professionnelle	5%	15%	0%	68%
09-Autre bac+3 et bac+4	4%	19%	0%	65%
10-Bac+5 sauf écoles	2%	11%	0%	68%
11-Ecoles (Ingé., commerce)	1%	7%	0%	66%
12-Doctorat	0%	2%	0%	48%
Ensemble	10%	10%	4%	74%

Champ : répondants (hors post-initiaux).

Note : la pondération associée aux répondants de l'enquête Génération n'est pas utilisée ici.

Source : Céreq-Depp. Rapprochement de l'enquête Génération 2017 à trois ans et des fichiers des apprenants FAERE.

Le questionnaire de l'enquête Génération comprend une question sur le redoublement au cours du collège (PS050). Elle n'est posée qu'aux jeunes inscrits dans une classe ou une formation de niveau au moins égal à la troisième et au plus égale au niveau bac+3 au cours de leur année scolaire de sortie de formation initiale, soit l'année scolaire 2016-2017. La réponse est, par nature, déclarative et la notion de « collège » peut être un peu plus large que celle que nous avons retenue pour le repérage des redoublements dans les données administratives, notamment du fait de l'exclusion des SEGPA. Ces limites étant rappelées et en prenant en compte les pondérations :

- 62,2% des personnes ayant indiqué avoir redoublé au collège dans Génération sont repérés comme tels à partir des fichiers des apprenants 2005 à 2016 (63,2% en données non pondérées) ;
- 0,8% des personnes ayant indiqué ne pas avoir redoublé au collège dans Génération sont repérés comme ayant redoublé à partir des fichiers des apprenants 2005 à 2016 (0,9% en données non pondérées) ;
- 2,7% des personnes hors du champs de la question dans Génération sont repérées comme ayant redoublé à partir des fichiers des apprenants 2005 à 2016.

Tableau 17 • Cohérence/Complémentarité de l'enquête et des données administratives sur le redoublement au collège, selon le plus haut diplôme atteint

	Sur tous les répondants à Génération 2017			Sur tous les répondants à Génération retrouvés au moins une fois dans les fichiers FAERE		% de repérés au moins une fois dans les fichiers de la DEPP
	% des redoublants dans l'enquête repérés comme tels dans FAERE	% des non redoublants dans l'enquête repérés comme redoublants dans FAERE	% des hors champs de la question à l'enquête repérés comme redoublants dans FAERE	% des redoublants dans l'enquête repérés comme tels dans FAERE	% des non redoublants dans l'enquête repérés comme redoublants dans FAERE	
01-Non diplômés	65,9%	2,0%	29,3%	73,3%	2,1%	96%
02-CAP-BEP-MC-Divers Niv.3	70,5%	1,8%	100,0%	79,8%	2,0%	94%
03-Bac pro-BT-Divers Niv.4	60,9%	0,6%	0,0%	81,9%	0,8%	75%
04-Bac technologique	68,6%	0,2%	4,6%	80,2%	0,2%	86%
05-Bac général	61,2%	0,4%	3,3%	77,4%	0,4%	85%
06-Bac+2 sauf santé/social	57,6%	0,5%	5,2%	76,6%	0,7%	73%
07-Bac+2 à +4 santé/social	13,6%	0,0%	0,0%	42,7%	0,0%	39%
08-Licence professionnelle	52,1%	0,2%	0,5%	76,7%	0,2%	68%
09-Autre bac+3 et bac+4	43,2%	0,4%	2,1%	55,2%	0,5%	65%
10-Bac+5 sauf écoles						68%
11-Ecoles (Ingé, commerce)						66%
12-Doctorat						48%
	62,2%	0,8%	2,7%	77,4%	1,0%	74,4%

Champ : répondants (hors post-initiaux).

Note : les données sont pondérées avec la pondération associée aux répondants de l'enquête Génération.

La question du redoublement en collège (PS050) n'est pas posée aux jeunes inscrits en 2016/2017 dans une formation de niveau bac+4 et plus, ou dans une classe antérieure à la troisième.

Source : Céreq-Depp. Rapprochement de l'enquête Génération 2017 à trois ans et des fichiers des apprenants FAERE.

4.4 Repérer les inscriptions en 3^{ème} et l'orientation en fin de 3^{ème} dans les données administratives et comparer avec ce que dit l'enquête

Le questionnaire de l'enquête Génération comprend une question sur la classe de troisième suivie (PS040). Elle n'est posée qu'aux jeunes inscrits dans une classe ou une formation de niveau au moins égal à la troisième et au plus égal au niveau bac+3 au cours de leur année scolaire de sortie de formation initiale, soit l'année scolaire 2016-2017. Par ailleurs, le repérage d'une inscription en troisième générale peut s'effectuer à partir des données administratives, à partir de la variable synthétique de classe présentée dans la partie 4.1.

En prenant en compte les pondérations de l'enquête, la comparaison globale des deux informations (tableau 18) montre que :

- 73 % des répondants à l'enquête déclarant avoir fait une troisième générale sont repérés comme tel avec les données administratives ;
- 17 % des répondants à l'enquête déclarant n'avoir fait aucune troisième, quel que soit son type, sont repérés comme ayant eu une inscription en troisième générale dans les données administratives ;
- 54 % des enquêtés à qui la question PS040 n'a pas été posée sont repérés comme ayant fait une troisième générale avec les données administratives.

Tableau 18 • Cohérence/Complémentarité de l'enquête et des données administratives sur l'inscription en classe de troisième

Nombre de troisièmes repérées dans FAERE	Réponse à la question PS040 sur la classe de troisième suivie pour les répondants à l'enquête génération 2017								
	4-Aucune 3e	1-3e générale	2-3e SEGPA, ULIS	3-3e prépa pro	5-3 ^e à l'étranger	6-Autres	Question non posée	Ensemble	
0	1 985	120 816	21 701	15 174	7 192	4 713	95 262	266 843	36%
1	394	311 259	2 680	30 315	588	3 186	110 795	459 216	62%
2	0	15 394	233	916	44	313	2 920	19 820	3%
3	0	9	0	0	0	0	112	121	0%
Total	2 378	447 478	24 615	46 405	7 824	8 212	209 088	746 000	100%
	0%	60%	3%	6%	1%	1%	28%	100%	

Part de ceux ayant au moins une troisième générale repérée dans FAERE :

	17%	73%	12%	67%	8%	43%	54%	64%
--	-----	-----	-----	-----	----	-----	-----	-----

Champ : répondants (hors post-initiaux).

Note : les données sont pondérées avec la pondération associée aux répondants de l'enquête Génération.

La question de l'inscription en troisième (PS040) n'est pas posée aux jeunes inscrits en 2016/2017 dans une formation de niveau bac+4 et plus, ou dans une classe antérieure à la troisième.

Source : Céreq-Depp. Rapprochement de l'enquête Génération 2017 à trois ans et des fichiers des apprenants FAERE.

Le tableau 19 permet de détailler ces indicateurs globaux selon le plus haut diplôme atteint :

- À tous les niveaux inférieurs ou égaux à bac+3, une petite proportion des répondants ne se voit pas poser la question. Par exemple, 6 % pour ceux qui ont un baccalauréat général comme plus haut diplôme ou 9 % chez les ceux qui ont au mieux une licence professionnelle.
- Ceux qui ont un baccalauréat général comme plus haut diplôme représentent la majorité (55 %) des répondants qui déclarent à l'enquête n'avoir fait aucune troisième, quel que soit son type, mais sont pourtant repérés comme ayant eu une inscription en troisième générale dans les données administratives. Les autres jeunes concernés n'ont aucun diplôme ou, au mieux, un CAP, BEP, ou bac pro.
- Inversement, en dehors des non diplômés, quel que soit le plus haut diplôme atteint, les données administratives mobilisées ne permettent pas de repérer une proportion importante des jeunes ayant fait une troisième générale. Par exemple, l'information n'est pas récupérée pour 22 % des jeunes ayant au mieux un CAP-BEP. C'est 30 % au niveau des bac professionnelle ou équivalent, 31 % pour les bacs+2, 32 % au niveau des licences professionnelles.

- Dans la très grande majorité des cas, les individus qui n'ont pu être repérés comme ayant fait une troisième générale à partir des fichiers FAERE alors qu'ils déclarent l'inverse à l'enquête, sont des individus qui n'ont pas pu être retrouvés du tout dans FAERE. Si on se limite aux répondants retrouvés au moins une fois dans FAERE, il reste toutefois 9 % des cas où le passage par une troisième générale est documenté par l'enquête sans l'être par les données administratives.

Tableau 19 • Cohérence/Complémentarité de l'enquête et des données administratives sur l'inscription en classe de troisième, selon le plus haut diplôme atteint

			Tous les répondants à l'enquête Génération 2017		Répondants retrouvés au moins une fois dans les fichiers FAERE 2005 à 2016	
	% des répondants déclarant avoir fait une 3e générale dans Génération	% de répondants de Génération à qui PS040 n'a pas été posée	% répondants ayant une 3e générale repérée dans FAERE parmi ceux déclarant n'avoir pas fait de 3e dans Génération	% répondants sans 3e générale repérée dans FAERE parmi ceux déclarant avoir fait une 3e générale dans Génération	% répondants ayant une 3e générale repérée dans FAERE parmi ceux déclarant n'avoir pas fait de 3e dans Génération	% répondants sans 3e générale repérée dans FAERE parmi ceux déclarant avoir fait une 3e générale dans Génération
01-Non diplômés	60%	11%	20%	14%	29%	9%
02-CAP-BEP-MC-Divers Niv.3	57%	1%	14%	22%	14%	14%
03-Bac pro-BT-Divers Niv.4	80%	1%	9%	30%	9%	9%
04-Bac technologique	95%	2%	0%	20%	s.o.	6%
05-Bac général	92%	6%	100%	17%	100%	5%
06-Bac+2 sauf santé/social	93%	3%	0%	31%	0%	10%
07-Bac+2 à +4 santé/social	96%	2%	s.o.	66%	s.o.	14%
08-Licence professionnelle	89%	9%	s.o.	32%	s.o.	9%
09-Autre bac+3 et bac+4	49%	49%	0%	29%	s.o.	12%
10-Bac+5 sauf écoles	4%	95%		56%		34%
11-Ecoles (Ingé, commerce)	1%	99%		75%		46%
12-Doctorat	0%	100%		100%		s.o.
Ensemble	60%	28%	17%	27%	20%	9%

Champ : répondants (hors post-initiaux).

Note : les données sont pondérées avec la pondération associée aux répondants de l'enquête Génération.

La question de l'inscription en troisième (PS040) n'est pas posée aux jeunes inscrits en 2016/2017 dans une formation de niveau bac+4 et plus, ou dans une classe antérieure à la troisième.

Source : Céreq-Depp. Rapprochement de l'enquête Génération 2017 à trois ans et des fichiers des apprenants FAERE.

Le questionnaire de l'enquête Génération comprend une question sur la classe suivie après la troisième (PS060). Elle est posée à tous les jeunes ayant déclaré avoir suivi une troisième à la question PS040 (troisième générale, professionnelle, SEGPA, à l'étranger ou autre). La question PS060 n'est donc pas posée aux jeunes inscrits dans une classe ou une formation de niveau inférieur à la troisième et supérieur à bac+3 au cours de leur année scolaire de sortie de formation initiale (2016-2017). Nous avons reconstitué partiellement cette information avec les données administratives en regardant, quand l'information est disponible, la classe d'inscription l'année suivant l'année d'inscription en troisième. Il faut donc que les deux années consécutives soient disponibles dans les fichiers de la DEPP. Comme précédemment, les classes sont identifiées à partir de la variable synthétique de classe présentée dans la partie 4.1. Les classes de troisième se limite donc aux troisièmes générales.

Le tableau 20 porte sur la cohérence entre l'enquête et les données administratives sur cette orientation post troisième. Il montre notamment que **dans une proportion importante des cas, l'orientation post troisième n'est pas reconstituée à partir des données administratives**. C'est notamment le cas pour 30 % des jeunes ayant fait une seconde générale ou professionnelle après leur troisième d'après l'enquête Génération. Cette proportion est de 29 % en cas d'orientation vers la seconde professionnelle et 46 % pour les orientations vers les premières années de CAP.

Quand l'orientation post troisième peut être reconstituée avec les données administratives, il y a dans la très grande majorité des cas une cohérence avec la déclaration faite à l'enquête, même si des petits écarts demeurent.

Ainsi :

- parmi les **répondants ayant déclaré avoir suivi une seconde générale après la troisième** pour lesquels une information est disponible dans FAERE, 97,2 % sont repérés avec une inscription en seconde générale après leur troisième dans FAERE ; 2,8 % sont dans un autre type de classe ;
- parmi les **répondants ayant déclaré avoir suivi une seconde professionnelle** après la troisième pour lesquels une information est disponible dans FAERE, 91,5 % sont repérés avec une inscription en lycée professionnel après leur troisième dans FAERE ; 6,9 % sont dans une classe de CAP et 1,5 % dans un autre type de classes ;
- parmi les **répondants ayant déclaré avoir suivi une première de CAP** après la troisième pour lesquels une information est disponible dans FAERE, 92,9 % sont repérés avec une inscription dans une classe de CAP après leur troisième dans FAERE ; 4,5 % avec une inscription en lycée professionnel et 2,6 % dans un autre type de classes ;

Tableau 20 • Cohérence/Complémentarité de l'enquête et des données administratives sur l'orientation après une troisième générale

Classe post troisième identifiée dans FAERE	Orientation après la troisième selon l'enquête Génération (réponse à la question PS060)									Total
	Seconde générale ou techno	Première année de CAP	Seconde pro	Pas de classe après la troisième	Seconde à l'étranger	1ère année de BEP	Autre	Question non posée	%	
06-SES/SEGPA	0	167	0	0	0	0	0	116	41%	283
07-Niv. Coll-autres	0	273	152	0	0	0	174	0	0%	600
08-2de générale	192 131	312	869	38	39	48	139	101 996	35%	295 572
09-1re générale	0	0	0	0	0	0	86	60	41%	146
13-Niv. Lycée-autres	2 348	352	863	42	0	0	269	373	9%	4 246
14-CAP	442	39 212	6 338	296	43	7	355	313	1%	47 006
15-CAP TERM	0	5	59	0	0	0	0	0	0%	64
20-BEP	190	48	1 522	0	0	1 036	0	684	20%	3 480
23-Bac pro	2 467	1 883	105 484	62	8	0	117	592	1%	110 614
98-Pas d'information FAERE	85 019	36 032	47 610	1 361	4 658	2 762	1 592	97 546	35%	276 578
99-troisième en 2016	0	0	0	0	0	0	0	7 412	100%	7 412
Total	282 595	78 286	162 897	1 798	4 747	3 853	2 732	209 092	28%	746 000
Proportion sans information FAERE	30%	46%	29%	76%	98%	72%	58%	50%		38%

Champ : répondants (hors post-initiaux).

Note : les données sont pondérées avec la pondération associée aux répondants de l'enquête Génération.

La question de l'inscription en troisième (PS060) n'est pas posée aux jeunes inscrits en 2016/2017 dans une formation de niveau bac+4 et plus, ou dans une classe antérieure à la troisième.

Source : Céreq-Depp. Rapprochement de l'enquête Génération 2017 à trois ans et des fichiers des apprenants FAERE.

4.5 Repérer les inscriptions et les réussites à certains diplômes dans les données administratives et comparer avec ce que dit l'enquête

Comme présenté dans la partie 1.1.b, les fichiers annuels des apprenants intègrent des informations issues des systèmes d'information relatifs aux examens passés que le ministère de l'Éducation nationale centralise, notamment par l'intermédiaire du système d'information OCEAN (Organisation des concours et examens académiques et nationaux OCEAN). Les examens intégrés sont les suivants : le diplôme national du brevet, le baccalauréat, le CAP, le BEP, le BP, les mentions complémentaires de niveau V et IV ainsi que certains diplômes supérieurs (BTS et diplômes comptables) à l'exception des spécialités agricoles.

Les fichiers relatifs aux examens ne comportent pas toujours l'INE, selon des proportions variables qui dépendent des années et des diplômes. Pour le brevet et le baccalauréat, l'INE est *a priori* bien renseigné alors que pour les CAP, la proportion d'INE renseigné peut parfois être faible sur la période couverte par nos travaux expérimentaux (la mise en œuvre de SYSCA, dans le cadre du déploiement de l'INE unique certifié devrait améliorer la situation à l'avenir). Comme expliqué dans l'encadré 3, la DEPP s'efforce de suppléer à l'absence de l'INE par un appariement sur profil parmi les élèves potentiellement inscrits aux examens (les « viviers »).

Grâce à ce rapprochement des fichiers des inscrits et des fichiers des examens, il est possible de savoir si l'élève est inscrit à l'examen, s'il a été effectivement présent ou non, ainsi que le résultat obtenu (mention).

Depuis 2013, pour chaque élève, deux inscriptions aux examens peuvent être enregistrées une même année dans le fichier des apprenants. Ces cas sont cependant extrêmement rares dans le champ des répondants à l'enquête Génération 2017. Ils concernent quasi systématiquement des jeunes inscrits en bac professionnel, également inscrits à un BEP ou CAP (tableau 21). Dans les exploitations présentées ensuite, sur le nombre d'inscriptions repérées au brevet, baccalauréat, CAP et BEP, nous avons retenu l'ensemble des informations disponibles, donc les deux inscriptions enregistrées dans les rares cas concernés.

Tableau 21 • Doubles Inscriptions à des examens repérées dans les fichiers FAERE 2013 à 2016 pour les répondants de Génération 2017

1er examen renseigné	Fichier 2016 2e examen renseigné			Fichier 2015 2e examen renseigné			Fichier 2014 2e examen renseigné		Fichier 2013 2e examen renseigné	
	Aucun	500 (CAP)	510 (BEP)	Aucun	500 (CAP)	510 (BEP)	Aucun	510 (BEP)	Aucun	510 (BEP)
01D	22			1			5		1	
01E	0			4			7		2	
400 (bac pro)	2 895	35	59	675	8	17	802	8	1 070	2
430 (bac techno)	379			241			621		71	
450 (brevet pro)	2			24			13		3	
470 (bac général)	154			247			471		553	
500 (CAP)	1 939	1		1 170		2	721		362	
510 (BEP)	190			1 600			481		588	
670 (DNB)	96			300			1 039		2 758	
Total	5 677			4 262			4 160		5 408	

Champ : répondants (hors post-initiaux).

Note : la pondération associée aux répondants de l'enquête Génération n'est pas utilisée ici.

Source : Céreq-Depp. Rapprochement de l'enquête Génération 2017 à trois ans et des fichiers des apprenants FAERE.

Nous n'avons pas réussi à trouver de nomenclature spécifique couvrant l'ensemble des modalités prises par la variable EXAM des fichiers FAERE. Nous avons donc défini nos agrégations en croisant les trois premières positions de la variable EXAM et la variable synthétique sur la classe d'inscription présentée dans la partie 4.1.

Tableau 22.a • Inscriptions aux examens repérées dans les fichiers de la DEPP pour les répondants de Génération 2017, selon la composante de la base de sondage

	Proportions de personne concernées parmi les répondants à l'enquête Génération 2017 repérés au moins une fois dans les fichiers FAERE 2005 à 2016						% de repérés au moins une fois dans les fichiers de la DEPP
	Brevet		Bac		CAP/BEP		
	1 fois	2 fois ou plus	1 fois	2 fois ou plus	1 fois	2 fois ou plus	
Sysca (second degré)	76%	5%	58%	6%	58%	7%	100%
Sifa (apprentissage)	69%	3%	46%	4%	42%	12%	100%
Universités	88%	1%	92%	5%	5%	0%	88%
Autres établissements	87%	0%	94%	2%	3%	0%	56%
DREES	90%	1%	93%	4%	5%	0%	20%
Culture	73%	0%	95%	0%	0%	0%	37%
DGAFP	87%	1%	96%	0%	0%	0%	41%
Agriculture, Sport, Cifre, Contrats de prof.	s.o.	s.o.	s.o.	s.o.	s.o.	s.o.	0%
Ensemble	80%	3%	71%	5%	32%	5%	74%

Champ : répondants (hors post-initiaux).

Note : la pondération associée aux répondants de l'enquête Génération n'est pas utilisée ici.

Source : Céreq-Depp. Rapprochement de l'enquête Génération 2017 à trois ans et des fichiers des apprenants FAERE.

Tableau 22.b • Inscriptions aux examens repérées dans les fichiers de la DEPP pour les répondants de Génération 2017, selon le plus haut diplôme atteint

	Proportion de personnes concernées parmi les répondants à l'enquête Génération 2017 repérés au moins une fois dans les fichiers FAERE 2005 à 2016						% de repérés au moins une fois dans les fichiers de la DEPP
	Brevet		Bac		CAP/BEP		
	1 fois	2 fois ou plus	1 fois	2 fois ou plus	1 fois	2 fois ou plus	
01-Non diplômés	68%	6%	11%	4%	30%	4%	96%
02-CAP-BEP-MC-Divers Niv.3	56%	5%	19%	2%	76%	18%	94%
03-Bac pro-BT-Divers Niv.4	82%	6%	80%	8%	76%	11%	75%
04-Bac technologique	91%	3%	88%	9%	4%	0%	86%
05-Bac général	92%	1%	84%	13%	4%	1%	85%
06-Bac+2 sauf santé/social	88%	2%	86%	6%	26%	1%	73%
07-Bac+2 à +4 santé/social	87%	0%	90%	5%	6%	0%	39%
08-Licence professionnelle	89%	1%	86%	4%	9%	0%	68%
09-Autre bac+3 et bac+4	89%	0%	90%	7%	4%	0%	65%
10-Bac+5 sauf écoles	85%	0%	92%	3%	1%	0%	68%
11-Ecoles (Ingé, commerce)	88%	0%	92%	1%	1%	0%	66%
12-Doctorat	22%	0%	92%	0%	0%	0%	48%
Ensemble	80%	3%	71%	5%	32%	5%	74%

Champ : répondants (hors post-initiaux).

Note : la pondération associée aux répondants de l'enquête Génération n'est pas utilisée ici.

Source : Céreq-Depp. Rapprochement de l'enquête Génération 2017 à trois ans et des fichiers des apprenants FAERE.

Les constats présentés dans les tableaux 22.a et 22.b suggèrent que **l'enrichissement par les données administratives sur les examens semble globalement plutôt bon, sans être exhaustif, dès lors que les répondants à l'enquête Génération 2017 sont retrouvés au moins une fois dans les fichiers de la DEPP sur la période 2005-2016**. En particulier :

- l'inscription au diplôme national du brevet est récupérée pour environ 9 jeunes sur 10 quel que soit le plus haut diplôme atteint, sauf pour les docteurs, pour lesquels la profondeur historique mobilisée ne permet souvent pas de remonter suffisamment loin pour la repérer. Pour les non-diplômés et, plus encore, les jeunes n'ayant pas plus qu'un CAP ou BEP comme diplôme, la proportion est aussi moindre, respectivement 74 % et 61 % ;
- des inscriptions au baccalauréat sont récupérées pour 88 % des répondants ayant un bac professionnel ou BT comme plus haut diplôme, 97 % pour les bacs technologiques et généraux. Elles sont aussi souvent disponibles pour les diplômés de l'enseignement supérieur ;
- des inscriptions au CAP/BEP sont récupérées pour 94 % des répondants ayant un CAP ou BEP comme plus haut diplôme, 87 % de ceux ayant obtenu au plus un bac professionnel ou BT, 34 % des non-diplômés.

Tableau 23 • Cohérence entre les informations issues de l'enquête et celles issues des données administratives sur la réussite aux examens pour les inscrits en année terminale de CAP/BEP et de baccalauréat en 2016/2017

	CAP/BEP	Bac pro	Bac techno	Bac général
Proportion d'appariés avec FAERE 2016 parmi les répondants de ce niveau de sortie	91%	87%	98%	98%
Proportion de répondants qui ne seraient pas repérés comme ayant réussi leur diplôme terminal avec la seule utilisation de FAERE parmi ceux déclarant l'avoir eu dans l'enquête :				
- parce qu'ils ne sont pas retrouvés dans FAERE	12%	14%	2%	3%
- parce qu'ils sont retrouvés dans FAERE sans être repérés comme ayant réussi le diplôme	15%	4%	0%	2%
Proportion des répondants repérés comme n'ayant pas réussi le diplôme dans FAERE (absents ou échecs) parmi ceux qui déclarent avoir eu leur diplôme terminal dans l'enquête	1%	1%	0%	2%
Proportion des répondants repérés comme ayant réussi le diplôme dans FAERE parmi ceux qui déclarent n'avoir pas eu leur diplôme terminal dans l'enquête	2%	1%	0%	0%
Proportion des répondants repérés comme inscrits à l'examen avec FAERE parmi ceux inscrits en année terminal du niveau d'après l'enquête	74%	81%	96%	94%
Même proportion en se limitant aux répondants retrouvés dans FAERE	83%	93%	97%	96%

Champ : répondants (hors post-initiaux).

Note : la pondération associée aux répondants de l'enquête Génération n'est pas utilisée ici.

Source : Céreq-Depp. Rapprochement de l'enquête Génération 2017 à trois ans et des fichiers des apprenants FAERE.

4.6 Examiner de façon plus précise le cas du baccalauréat

L'enquête Génération permet de repérer les sortants au niveau baccalauréat qui ont obtenu leur diplôme. Elle demande aussi plus largement à l'ensemble des répondants potentiellement concernés, quel que soit leur niveau de sortie, s'ils ont obtenu ce diplôme et si oui, l'année d'obtention et la mention obtenue. Il est possible de comparer les réponses à ces questions et les informations administratives de la DEPP en mobilisant l'ensemble des fichiers FAERE récupérés. Les comparaisons qui suivent sont faites sans utiliser la pondération de l'enquête.

Concernant le fait d'avoir le baccalauréat (tableau 24) :

- Parmi les répondants à l'enquête retrouvés au moins une fois dans les fichiers de la DEPP sur la période 2005-2016, 93 % de ceux qui déclarent avoir le baccalauréat à l'enquête sont repérés comme tels à partir des données administratives. Corrélativement, 7 % des répondants qui déclarent avoir le baccalauréat à l'enquête n'ont pas de réussite à l'examen repérée dans les données administratives.
- Si l'on regarde tous les répondants à l'enquête, y compris ceux qui ne sont jamais retrouvés dans les fichiers de la DEPP, 34 % des répondants déclarant avoir le baccalauréat à l'enquête ne peuvent être identifiées comme tels avec les données administratives mobilisées.
- Inversement, 0,3 % de l'ensemble des répondants sont identifiés comme ayant réussi le baccalauréat avec les données administratives alors qu'ils ne le déclarent pas à l'enquête. Ils représentent 1,3 % des répondants considérés comme non bacheliers par l'enquête.

Tableau 24 • Avoir le bac dans Génération et être identifié comme ayant réussi l'examen dans l'un des fichier FAERE 2005 à 2016

	D'après les fichiers de la DEPP 2005 à 2016, réussite au moins une fois au								Total
	bac général		bac techno		bac pro		bac (tous types)		
	Non	Oui	Non	Oui	Non	Oui	Non	Oui	
Répondants à l'enquête Génération non retrouvés dans les fichiers FAERE 2005 à 2006									
Pas de bac selon l'enquête	467	0	467	0	467	0	467	0	467
Bac selon l'enquête	5 985	0	5 985	0	5 985	0	5 985	0	5 985
Répondants à l'enquête Génération retrouvés dans les fichiers FAERE 2005 à 2006									
Pas de bac selon l'enquête	4 336	15	4 342	9	4 312	39	4 289	62	4 351
Bac selon l'enquête	7 759	6 602	12 413	1 948	9 493	4 868	972	13 389	14 361
Ensemble	18 547	6 617	23 207	1 957	20 257	4 907	11 713	13 451	18 712

Champ : répondants (hors post-initiaux).

Note : la pondération associée aux répondants de l'enquête Génération n'est pas utilisée ici.

Source : Céreq-Depp. Rapprochement de l'enquête Génération 2017 à trois ans et des fichiers des apprenants FAERE.

Pour les 13 389 répondants à l'enquête qui déclarent avoir eu le baccalauréat et qui sont repérés comme tels dans les données administratives mobilisées, il est possible de regarder si les informations sont cohérentes quant au type de baccalauréat, à l'année d'obtention et à la mention obtenue. Si l'enquête n'enregistre qu'un seul baccalauréat, les fichiers administratifs peuvent en repérer plusieurs. Seuls 29 répondants sont dans ce cas. L'examen de la cohérence des informations entre les deux sources se limite donc aux 13 360 répondants bacheliers n'ayant qu'un baccalauréat repéré dans les données administratives.

Dans le champ retenu, concernant l'année du baccalauréat :

- l'année est identique si l'on utilise l'une ou l'autre source dans 93,4 % des cas ;
- l'écart est d'une année dans 5,4 % des cas (un an plus tôt avec l'enquête par rapport aux données administratives dans 3,0 % des cas ; situation inverse dans 2,4 % des cas) ;
- l'écart est plus important dans 1,3 % des cas.

Concernant le type de baccalauréat obtenu (tableau 25),

- 99,8 % des répondants ayant réussi un baccalauréat général d'après les données administratives déclarent avoir un baccalauréat général à l'enquête Génération ;
- 99,4 % des répondants ayant réussi un baccalauréat technologique d'après les données administratives déclarent avoir un baccalauréat technologique à l'enquête Génération ;
- 79,3 % des répondants ayant réussi un baccalauréat professionnel d'après les données administratives déclarent avoir un baccalauréat professionnel à l'enquête Génération, 14,6 % déclarent avoir un baccalauréat technologique et 9,2 %, un baccalauréat général.

Nous n'avons pu déterminer les raisons de cet écart observé pour les baccalauréats professionnels. Dans leur cas, la divergence entre la donnée administrative et la donnée d'enquête est plus forte pour la collecte par internet que pour la collecte par téléphone, puisque les taux de cohérence sont respectivement de 66 % sur Internet et 85 % par téléphone. 60 % des cas de ces divergences concernent des jeunes sortis de BTS ou de licences professionnelles³⁴ et 29 % des jeunes sortis de licences générales.

Tableau 25 • Cohérence du type de baccalauréat déclaré à l'enquête avec celui repéré dans les données administratives

Type de baccalauréat selon l'enquête Génération	Réussite au moins une fois au bac général dans FAERE		Réussite au moins une fois au bac techno dans FAERE		Réussite au moins une fois au bac pro dans FAERE		Total
	Non	Oui	Non	Oui	Non	Oui	
Bac général	451	6 579	7 023	7	6 586	444	7 030
Bac technologique	2 478	3	562	1 919	1 922	559	2 481
Bac professionnel	3 839	8	3 843	4	12	3 835	3 847
Autre bac (étranger notamment)	1	1	2	0	1	1	2
Total	6 769	6 591	11 430	1 930	8 521	4 839	13 360

Champ : répondants (hors post-initiaux) déclarant avoir eu le baccalauréat et repéré une seule fois comme tel dans les fichiers FAERE 2005 à 2016.

Note : la pondération associée aux répondants de l'enquête Génération n'est pas utilisée ici.

Source : Céreq-Depp. Rapprochement de l'enquête Génération 2017 à trois ans et des fichiers des apprenants FAERE.

Dans le tableau 26, la cohérence entre les deux sources n'est plus analysée par rapport à l'information administrative, mais du point de vue inverse, par rapport à l'enquête Génération. De ce point de vue, il y a plus de cohérence pour les titulaires d'un baccalauréat professionnel que pour les autres. On retrouve le fait que ceux ayant des licences professionnelles ou des bacs+2 hors santé/social comme plus haut diplôme, sont ceux pour qui il y a le plus de cas de divergences entre les données administratives et l'enquête.

Le tableau 27 compare la mention au baccalauréat enregistrée dans les données administratives et celles figurant dans l'enquête. Il y a convergence dans 92 % des cas. En cas de divergence, il est six fois plus fréquent d'avoir une meilleure mention dans l'enquête que dans les données administratives que le cas inverse, ce qui pourrait suggérer l'existence d'un léger biais déclaratif.

³⁴ Analyse selon la variable de niveau de sortie.

Tableau 26 • Cohérence du type de baccalauréat déclaré à l'enquête avec celui repéré dans les données administratives

	Non	Oui	Taux de cohérence
Selon le type de baccalauréat dans l'enquête Génération			
Baccalauréat général	821	6209	88,3%
Baccalauréat technologique	722	1759	70,9%
Baccalauréat professionnel	256	3591	93,3%
Autres	2	0	n.s.
Selon la composante de la base de sondage			
Sysca (second degré)	332	3764	91,9%
Sifa (apprentissage)	234	1158	83,2%
Universités	1028	4896	82,6%
Autres établissements	173	1476	89,5%
DREES	23	119	83,8%
Culture	4	48	n.s.
DGAFP	7	98	93,3%
Agriculture, Sport, Cifre, Contrats de prof.			s.o.
Selon le plus haut diplôme atteint			
03-Bac pro-BT-Divers Niv.4	168	3002	94,7%
04-Bac technologique	115	671	85,4%
05-Bac général	133	650	83,0%
06-Bac+2 sauf santé/social	467	1486	76,1%
07-Bac+2 à +4 santé/social	40	189	82,5%
08-Licence professionnelle	377	511	57,5%
09-Autre bac+3 et bac+4	265	767	74,3%
10-Bac+5 sauf écoles	153	2703	94,6%
11-Ecoles (Ingé, commerce)	65	1366	95,5%
12-Doctorat	18	214	92,2%
Ensemble	1801	11559	86,5%

Champ : répondants (hors post-initiaux) déclarant avoir eu le baccalauréat et repéré une seule fois comme tel dans les fichiers FAERE 2005 à 2016. Note : la pondération associée aux répondants de l'enquête Génération n'est pas utilisée ici.

Source : Céreq-Depp. Rapprochement de l'enquête Génération 2017 à trois ans et des fichiers des apprenants FAERE.

Tableau 27 • Cohérence de la mention au baccalauréat entre l'enquête et les données administratives

Mention du bac dans l'enquête	Mention du bac dans les données administratives				Total
	Passable	Assez bien	Bien	Très bien	
Passable	5 640	58	9	1	5 708
Assez Bien	523	3 900	72	3	4498
Bien	58	241	2 017	4	2 320
Très bien	5	12	38	713	768
Etranger	11	5	1	0	17
Non renseigné	32	17	0	0	49
Total	6 269	4 233	2 137	721	13 360

Champ : répondants (hors post-initiaux) déclarant avoir eu le baccalauréat et repéré une seule fois comme tel dans les fichiers FAERE 2005 à 2016. Note : la pondération associée aux répondants de l'enquête Génération n'est pas utilisée ici.

Source : Céreq-Depp. Rapprochement de l'enquête Génération 2017 à trois ans et des fichiers des apprenants FAERE.

5. Quelques informations complémentaires issues des données administratives

5.1 La nationalité de l'élève est renseignée de façon hétérogène selon la source

Les données administratives, comme l'enquête, proposent une information sur la nationalité de l'élève. On ne peut cependant pas les comparer directement et exhaustivement car elles ne sont pas homogènes.

Pour les données administratives, les deux sources qui alimentent les fichiers des apprenants ne proposent pas le même degré de détail. **Dans la source SYSCA, la nationalité est disponible de façon détaillée alors que SIFA ne propose que la distinction entre la nationalité française ou une nationalité étrangère.**

Pour l'enquête Génération 2017, la question sur la nationalité des répondants (SD040) n'est posée que dans l'objectif de l'utiliser ensuite avec le pays de naissance pour identifier les jeunes immigrés, afin de pouvoir comparer leurs trajectoires d'insertion à celles des autres. Un immigré étant, selon la définition de l'Insee, une personne née étrangère à l'étranger, la SD040 **n'est posée qu'aux jeunes qui déclarent être nés dans un autre pays que la France** (et pour lesquels l'information n'était pas disponible dans la base de sondage). Le pays de naissance est renseigné de façon précise tandis que la nationalité n'est collectée que de façon agrégée : française de naissance, française par acquisition, étrangère, double nationalité. Dans les fichiers de résultats, la nationalité est encore plus agrégée pour n'être proposée que de façon binaire (nationalité française ou étrangère). Pour les jeunes nées en France, la nationalité française est présumée. Il n'est pas possible de savoir pour qui l'information était disponible dans la base de sondage.

Pour les répondants à l'enquête Génération, nous avons comparé la nationalité mentionnée dans le fichier annuel des apprenants 2016, qui correspond à leur année de sortie du système éducatif, à celle enregistrée dans l'enquête Génération, en se limitant à l'alternative « française/étrangère » (tableau 28). Dans 55% des cas, l'information n'est pas disponible dans le fichier des apprenants 2016. **Quand l'information est disponible, il y a concordance sur l'information binaire « Français/Etranger » dans 96% des cas. Quand il y a divergence, dans trois quarts des cas, une nationalité étrangère est enregistrée dans l'enquête Génération alors que les données administratives mentionnent une nationalité française.**

Tableau 28 • Comparaison de la nationalité déclarée à l'enquête avec celle repérée dans les données administratives

Selon le fichier des apprenants 2016	Selon l'enquête Génération	
	Nationalité française	Nationalité étrangère
Nationalité française	10 389	323
Nationalité étrangère	92	529
Non disponible	12 702	1 129

Champ : répondants (hors post-initiaux).

Note : la pondération associée aux répondants de l'enquête Génération n'est pas utilisée ici.

Source : Céreq-Depp. Rapprochement de l'enquête Génération 2017 à trois ans et des fichiers des apprenants FAERE.

Cette comparaison n'utilise que le fichier administratif relatif à l'année de sortie. Dans le cas où l'information fait défaut cette année-là, on aurait pu chercher à la récupérer dans les fichiers plus anciens. Comme les trois quarts des répondants sont retrouvés au moins une fois dans les fichiers administratifs mobilisés, on devrait pouvoir récupérer une information pour la moitié des cas « non disponibles » dans le tableau 28.

Auparavant, nous avons cherché à voir si, pour les individus présents dans deux fichiers administratifs, l'information sur la nationalité est stable (tableau 29). D'une année sur l'autre, elle l'est dans environ 99% des

cas. En revanche, quand l'écart est plus important, on peut rapidement atteindre un taux de divergence de l'ordre de 4% entre deux fichiers administratifs, soit le même ordre de grandeur que celui observé en comparant l'enquête Génération aux données administratives 2016. En cas de divergence entre deux fichiers administratifs, les cas de bascules d'une nationalité étrangère vers une nationalité française sont du même ordre de grandeur que les cas de bascules inverses, alors que nous attendions plutôt à une bascule progressive vers la nationalité française. Quand l'analyse n'est faite que sur les données administratives issues de SYSCA, on observe davantage de bascules vers la nationalité française que de cas inverses.

Tableau 29 • Stabilité de la nationalité enregistrée dans les données administratives au cours du temps

Comparaison d'une année sur l'autre						
Année de fin	2016	2015	2014	2013	2012	2011
Année de début	2015	2014	2013	2012	2011	2010
Taux de stabilité	98,8%	98,6%	99,0%	99,2%	99,3%	99,4%
Etrangère-> Française	0,3%	0,4%	0,3%	0,3%	0,3%	0,3%
Française -> Etrangère	0,2%	0,3%	0,4%	0,4%	0,4%	0,3%
Comparaison de 2016 par rapport à une année plus ancienne						
Année de fin	2016	2016	2016	2016	2016	2016
Année de début	2015	2014	2013	2012	2011	2010
Taux de stabilité	98,8%	97,6%	97,1%	96,5%	96,3%	96,3%
Etrangère-> Française	0,3%	0,6%	0,8%	1,1%	1,1%	1,2%
Française -> Etrangère	0,2%	0,5%	0,8%	1,0%	1,1%	0,9%

Champ : les présents-présents dans les extraits de fichiers des apprenants comparés (restreints à l'échantillon de l'enquête).

Source : Céreq-Depp. Rapprochement de l'enquête Génération 2017 à trois ans et des fichiers des apprenants FAERE.

5.2 La catégorie sociale des parents n'est disponible que partiellement

Dans l'enquête Génération 2017, la catégorie sociale de chacun des deux parents est collectée. Elle porte sur la situation au moment de la fin des études du répondant. Pour chacun d'eux :

- une première question demande si le parent était en emploi au moment de la fin des études. L'absence du parent ou son décès font partie des réponses possibles ;
- si le parent n'était pas en emploi en moment de la fin des études, une seconde question demande s'il était retraité, au chômage, au foyer après avoir déjà travaillé, au foyer sans avoir jamais travaillé, ou en formation ;
- pour ceux qui étaient en emploi ou qui avaient déjà travaillé, une troisième question demande la catégorie socioprofessionnelle du dernier emploi occupé, parmi six modalités : « agriculteur », « artisan, commerçant, chef d'entreprise », « cadre, ingénieur, profession libérale », « professeur », « technicien, agent de maîtrise, VRP, profession intermédiaire », « employé de bureau, de commerce ou personnel de services », « ouvrier ».
- une dernière question demande si le parent était fonctionnaire.

Dans les données administratives mobilisées, les informations disponibles et leur complétude dépend de la source d'origine :

- **La source SIFA n'enregistre la PCS que d'une seule « personne responsable », sans que le lien de parenté avec l'élève ne soit précisé.** Selon la documentation de la DEPP, le **taux d'observations non renseignées est de l'ordre de 50%**, voire 60% pour les apprentis de l'enseignement supérieur. Dans le fichier 2016 mobilisé pour nos travaux, ce taux est aussi de 60% ;
- **La source SYSCA enregistre la PCS de deux « personnes responsables ». Une variable permet de préciser leurs liens de parenté avec l'élève.** Dans le fichier 2016 mobilisé pour nos travaux, le **taux d'observations non renseignées est de 8% pour la première personne responsable et de 41% pour la seconde.** La première personne « responsable » est la mère dans 47% des cas et le

père dans 46% des cas³⁵. La deuxième personne « responsable » est la mère dans 24% des cas et le père dans 39% des cas³⁶. Au total, la PCS de la mère est donc connue pour 71% des élèves et celle du père pour 84%. La codification est effectuée selon une nomenclature en 32 postes propre à l'Éducation nationale, à partir de celle de l'Insee en 42 postes. Notons également que l'information n'est pas disponible pour la rentrée scolaire 2008, année de polémique autour de l'identifiant national de l'élève.

Dans ce contexte, il est difficile et peu efficient de chercher à comparer l'information dans l'enquête Génération et les données administratives. Il est toutefois intéressant de souligner que **cette variable vient illustrer plusieurs limites posées par l'usage de données administratives à la place de données d'enquête** :

- la dépendance de l'information collectée aux pratiques administratives des acteurs (PCS des personnes « responsables », plutôt que des « parents ») ;
- l'absence d'homogénéité entre acteurs dans la façon de collecter une information de même nature (un seul responsable dans SIFA, deux dans SYSCA) ;
- une rigueur moindre dans la collecte d'informations non essentielles pour les opérations de gestion de l'acteur (d'où les taux de non renseignés considérables dans SIFA) ;
- les risques de perturbations externes (absence d'information en 2008).

5.3 Les département de résidence en 6^{ème} et au baccalauréat

L'enquête Génération collecte le département de résidence pour les répondants ayant déclaré avoir fait leur classe de sixième en France. 23 811 répondants sont dans ce cas. Le département de résidence en 6^e est connu dans les données administratives mobilisées pour 12 199 d'entre eux, soit 51,2%. Pour ces répondants pour lesquels l'information est disponible dans les deux sources, le département ne diffère entre les deux sources que pour 206 répondants, soit 1,7% des cas comparés.

L'enquête Génération collecte également le département de résidence au baccalauréat. Plus précisément, la question est posée par référence à l'année d'obtention du baccalauréat ou, à défaut et quand il y a lieu, par référence à l'année de la classe de terminale. Ce département de résidence a été comparé au département de résidence lors de la dernière inscription connue au diplôme du baccalauréat dans les données administratives mobilisées. On dispose de l'information dans Génération pour 20 346 répondants. L'information est également disponible dans les données administratives pour 13 308 d'entre eux, soit 65,4%. Les deux informations ne diffèrent que dans 309 cas, soit 2,3% des cas comparés.

³⁵ % calculé sur tous, y compris les non renseignés.

³⁶ Idem.

5.4 Autres informations complémentaires illustratives

Les tableaux présentés donnent des exemples d'informations complémentaires issues des données administratives mobilisés, qui pourraient, à termes, enrichir l'enquête Génération.

Tableau 30a • Les répondants repérés comme ayant été boursier au moins une fois au collège (y compris)

	Effectif	% de répondants jamais repérés dans une classe concernée (collège)	% de boursiers parmi ceux repérés au moins une fois dans une classe concernée
01-Non diplômés	1 779	11%	58%
02-CAP-BEP-MC-Divers Niv.3	978	18%	51%
03-Bac pro-BT-Divers Niv.4	4 955	31%	32%
04-Bac technologique	951	19%	33%
05-Bac général	959	20%	25%
06-Bac+2 sauf santé/social	2 948	33%	20%
07-Bac+2 à +4 santé/social	623	66%	6%
08-Licence professionnelle	1 463	38%	13%
09-Autre bac+3 et bac+4	1 649	42%	14%
10-Bac+5 sauf écoles	4 431	42%	9%
11-Ecoles (Ingé, commerce)	2 350	41%	5%
12-Doctorat	525	90%	0%

Champ : répondants (hors post-initiaux).

Note : la pondération associée aux répondants de l'enquête Génération n'est pas utilisée ici.

Source : Céreq-Depp. Rapprochement de l'enquête Génération 2017 à trois ans et des fichiers des apprenants FAERE.

Tableau 30b • Les répondants repérés comme ayant été boursier au moins une fois au lycée (classes générales ou technologiques)

	Effectif	% de répondants jamais repérés dans une classe concernée (lycée)	% de boursiers parmi ceux repérés au moins une fois dans une classe concernée
01-Non diplômés	1 779	80%	7%
02-CAP-BEP-MC-Divers Niv.3	978	96%	1%
03-Bac pro-BT-Divers Niv.4	4 955	92%	2%
04-Bac technologique	951	17%	25%
05-Bac général	959	17%	21%
06-Bac+2 sauf santé/social	2 948	50%	10%
07-Bac+2 à +4 santé/social	623	64%	4%
08-Licence professionnelle	1 463	44%	10%
09-Autre bac+3 et bac+4	1 649	38%	14%
10-Bac+5 sauf écoles	4 431	35%	10%
11-Ecoles (Ingé, commerce)	2 350	39%	5%
12-Doctorat	525	56%	3%

Champ : répondants (hors post-initiaux).

Note : la pondération associée aux répondants de l'enquête Génération n'est pas utilisée ici.

Source : Céreq-Depp. Rapprochement de l'enquête Génération 2017 à trois ans et des fichiers des apprenants FAERE.

Tableau 31 • Nombre d'années d'inscriptions dans le secteur public et dans le secteur privé repérées dans les fichiers des apprenants 2005 à 2016

	Nombre moyen d'années d'inscriptions		
	Ensemble	dans le secteur privé	dans le secteur public
Selon la composante de la base de sondage			
Sysca (second degré)	7,1	0,9	6,2
Sifa (apprentissage)	7,2	2,7	4,5
Universités	6,7	1,2	5,5
Autres établissements	6,6	1,9	4,6
DREES	6,7	2,4	4,3
Culture	5,0	1,1	3,9
DGAFP	7,0	1,5	5,5
Agriculture, Sport, Cifre, Contrats de prof.	n.d.	n.d.	n.d.
Selon le plus haut diplôme atteint			
01-Non diplômés	5,8	0,6	5,2
02-CAP-BEP-MC-Divers Niv.3	6,4	1,3	5,2
03-Bac pro-BT-Divers Niv.4	7,4	1,5	5,9
04-Bac technologique	7,5	1,1	6,5
05-Bac général	7,1	1,2	5,9
06-Bac+2 sauf santé/social	8,3	1,8	6,5
07-Bac+2 à +4 santé/social	7,0	2,3	4,7
08-Licence professionnelle	7,8	1,5	6,3
09-Autre bac+3 et bac+4	6,6	1,2	5,3
10-Bac+5 sauf écoles	6,1	1,3	4,8
11-Ecoles (Ingé, commerce)	7,3	2,2	5,2
12-Doctorat	2,9	0,6	2,3
Ensemble	6,9	1,4	5,5

Champ : répondants (hors post-initiaux).

Note : la pondération associée aux répondants de l'enquête Génération n'est pas utilisée ici.

Source : Céreq-Depp. Rapprochement de l'enquête Génération 2017 à trois ans et des fichiers des apprenants FAERE.

6. La non-réponse

La « non-réponse » de l'enquête Génération englobe plusieurs aspects : la capacité à retrouver des coordonnées des personnes, l'appartenance de la personne au champ de l'enquête et l'acceptation par la personne contactée à répondre à l'enquête.

Dans la plupart des enquêtes de la Statistique publique, la « non-réponse » concerne essentiellement le dernier aspect. Dans les enquêtes « ménages » usuelles, on échantillonne en effet des logements, dont l'adresse est donc connue, puis, au sein de chaque logement, on échantillonne une personne parmi celles qui y résident.

Dans le cas de l'enquête Génération, les deux premiers aspects sont en revanche des composantes importantes. Les coordonnées disponibles dans la base de sondage sont datées d'environ trois ans avant l'enquête, année de la sortie de formation initiale. Quant à l'appartenance au champ de l'enquête, elle ne peut être validée que par un questionnement filtre au premier contact, notamment pour vérifier l'absence de poursuite d'études et l'absence d'interruption des études avant l'année de sortie, puisque l'enquête porte sur les sorties de formation initiale.

L'appariement expérimental devait permettre de voir si les données administratives permettent d'améliorer le traitement de la non-réponse globale. Nous n'avons finalement étudié ce sujet que de façon rapide car les travaux réguliers du Céreq sur le sujet ont montré l'importance des variables relatives au niveau de sortie et à la nature de la formation suivie. L'examen de l'apport d'informations supplémentaires doit donc pouvoir s'articuler à cette approche. Or, comme nous l'avons vu, les taux de couverture ne sont très bons que pour un nombre limité de niveaux de sortie, issus de composantes de la base de sondage alimentées par la DEPP. Or, sur ces champs, le Céreq dispose déjà d'informations analogues à celles contenues dans le fichier des apprenants de l'année scolaire de sortie pour traiter la non-réponse. Nous avons donc examiné s'il y avait des variations significatives du taux de réponse selon des variables de parcours construites à partir de l'ensemble des fichiers mobilisés, en particulier l'existence d'un redoublement au collège, un redoublement au lycée ou l'obtention du diplôme national du brevet. Le taux de réponse est un peu plus élevé en cas de redoublement et il est plus faible en cas de non obtention du diplôme national du brevet.

Une analyse logistique montre que ces informations sont cependant secondaires par rapport à d'autres, notamment la classe de sortie, la plus déterminante, puis le fait d'être boursier l'année de sortie et la CS de la première personne responsable l'année de sortie – ces deux informations ne sont pas utilisées dans le traitement de la non-réponse par le Céreq, notamment parce qu'elles ne sont disponibles que de façon parcellaire (notamment dans le segment SIFA). Le Céreq utilise cependant une information, a priori en partie corrélée : l'indicatrice de résidence dans un quartier de politique de la ville. Faute de disposer de cette information, nous n'avons pas pu prendre en compte cette information.

ANNEXE 1A. Nomenclature transversale d'identification des classes

Dans les fichiers des apprenants de la DEPP, la variable `FORTRAMS_7` propose une nomenclature d'identification des classes construite de façon à être transversale aux deux systèmes d'information qui les alimentent : SYSCA et SIFA.

<code>fortrams_7</code>	<code>libelle_court</code>	<code>libelle_long</code>
2000000	1ER CYCLE	2ND DEGRE - PREMIER CYCLE
2110000	1ER CYCLE GENE	2ND DEGRE - 1ER CYCLE GENERAL
2111141	6EMES	6EMES
2111242	5EMES	5EMES
2111343	4E AIDE-SOUTIEN	4EME D'AIDE ET DE SOUTIEN
2111443	4E GENE NC A&S	4EME GENE (NC 4E D'AIDE ET DE SOUTIEN)
2111544	3EMES GENE	3EMES GENERALES
2111644	3EME INSERTION	3EME D'INSERTION
2112141	6EME SEGPA	6EME SES/SEGPA
2112242	5EME SEGPA	5EME SES/SEGPA
2112343	4EME SEGPA	4EME SES/SEGPA
2112444	3EME SEGPA	3EME SES/SEGPA
2181111	DISPO RELAIS	DISPOSITIF RELAIS
2181211	ULIS	UNITES LOCALISEES INCLUSION SCO.(EX UPI)
2211143	4EME TECHNO	4EME TECHNOLOGIQUE
2211244	3EME TECHNO	3EME TECHNOLOGIQUE
2381111	CPA	CPA
2381211	CLIPA	CLIPA
2381311	APP JUNIOR	FORMATION D'APPRENTI JUNIOR
2381411	DIMA	DISPO D'INITIATION METIERS EN ALRTENANC
3000000	2ND CYCLE	2ND DEGRE - 2ND CYCLE GENE/TECHNO/PRO
3081111	MLDS-ITHAQUE	MDLS - ITINERAIRE D'ACCES A LA QUALIF
3081211	MLDS-SIO	MDLS - SESSION D'INFO ET D'ORIENTATION
3081311	MLDS-CIPPA	MDLS -CYCLE INSERTION PRO PAR ALTERNANCE
3081411	MLDS-MODAL NIV-	MDLS - MODAL NIV V
3081511	MLDS-PARCOURS	MDLS - ACCOMPAGNEMENT PARCOURS FORMATION
3082111	MLDS-MODAL NIV-	MDSL - MODAL NIV IV
3110000	2ND CYCLE GENE	2ND DEGRE - 2ND CYCLE GENERAL
3111131	2NDE GT	2NDE GT DE DETERMINATION
3112132	1ERE GENERALE	1ERE GENERALE
3112232	1ERE S	1ERE GENE SCIENTIFQUE
3112332	1ERE L	1ERE GENE LITTERAIRE
3112432	1ERE ES	1ERE GENE ECONOMIQUE ET SOCIALE
3113133	TERM GENERALE	TERMINALE GENERALE
3113233	TERM S	TERMINALE GENE SCIENTIFQUE
3113333	TERM L	TERMINALE GENE LITTERAIRE
3113433	TERM ES	TERMINALE GENE ECONOMIQUE ET SOCIALE
3211031	2NDE TECH BT/A	2NDE TECHNO, BT BTA, 2NDE SPECIALE
3211131	2NDE TMD HOT	2NDE TECHNO SPECIFIQUE TMD HOTEL.
3211231	2NDE BT	2NDE BT
3212032	1ERE TECH BT/A	1ERE TECHNO, BT BTA
3212132	1ERE STI	1ERE SCIENCES & TECHNO INDUSTRIELLES
3212232	1ERE STMG	1ERE SCIENCES & TECHNO GESTION STG/STT
3212332	1ERE STL	1ERE SCIENCES & TECHNO DE LABORATOIRE
3212432	1E ST2S/SMS	1ERE SC.& TECHNO SANTE SOCIAL ST2S/SMS

3212532	1ERE TMD HOT	1ERE TECHNO MUSIQUE-DANSE, HOTELLERIE
3212632	1ERE BTN AGRI	1ERE TECHNO AGRICOLE - STAE STPA STAV
3212732	1ERE STD2A	1ERE SCIENCES & TECHNO ARTS APPLIQUES
3212832	1ERE STI2D	1ERE STI DEVELOPPEMENT DURABLE
3213033	TERM TECH BT/A	TERMINALE TECHNO, BT BTA
3213133	TERM STI	TERMINALE SCIENCES&TECHNO INDUSTRIELLES
3213233	TERM STMG	TERMINALE SCIENCES&TECH GESTION STG/STT
3213333	TERM STL	TERMINALE SCIENCES & TECHNO LABORATOIRE
3213433	TERM SMS/ST2S	TERMINALE SC.& TECHNO SANTE SOCIAL
3213533	TERM TMD HOT	TERMINALE TECHNO MUSIQUE-DANSE, HOTEL.
3213633	TERM BTN AGRI	TERMINALE TECHNO AGRI -STAE STPA STAV
3213733	TERM STD2A	TERMINALE TECHNO STD2A
3213833	TERM STI2D	TERMINALE TECHNO STI2D
3301111	DIPPRO N5 1 AN	DIPLOME PROFESSIONNEL DE NIV V EN 1 AN
3301121	1/2 DIPPRO N5	DIPLOME PRO DE NIV V EN 2 ANS : 1ERE AN
3301122	2/2 DIPPRO N5	DIPLOME PRO DE NIV V EN 2 ANS : 2EME AN
3301131	1/3 DIPPRO N5	DIPLOME PRO DE NIV V EN 3 ANS : 1ERE AN
3301132	2/3 DIPPRO N5	DIPLOME PRO DE NIV V EN 3 ANS : 2EME AN
3301133	3/3 DIPPRO N5	DIPLOME PRO DE NIV V EN 3 ANS : 3EME AN
3309100	PDIV PREBAC	PREPARATION DIVERSE PRE-BAC
3309121	1/2 DIV PRE-BAC	PREPARATION DIVERSE PRE-BAC : 1E ANNEE
3309122	1/2 DIV PRE-BAC	PREPARATION DIVERSE PRE-BAC : 2E ANNEE
3310000	2ND CYCLE PRO	2ND DEGRE - 2ND CYCLE PROFESSIONNEL
3311011	CAP-CAPA 1 AN	CAP-CAPA EN 1 AN
3311021	1/2 CAP-CAPA	CAP-CAPA EN 2 ANS : 1ERE ANNEE
3311022	2/2 CAP-CAPA	CAP-CAPA EN 2 ANS : 2EME ANNEE
3311031	1/3 CAP-CAPA	CAP-CAPA EN 3 ANS : 1ERE ANNEE
3311032	2/3 CAP-CAPA	CAP-CAPA EN 3 ANS : 2EME ANNEE
3311033	3/3 CAP-CAPA	CAP-CAPA EN 3 ANS : 3EME ANNEE
3311321	SEGPA - 1/2 CAP	SES/SEGPA - CAP EN 2 ANS : 1ERE ANNEE
3311322	SEGPA - 2/2 CAP	SES/SEGPA - CAP EN 2 ANS : 2EME ANNEE
3312011	BEP-BEPA 1 AN	BEP-BEPA EN 1 AN
3312021	1/2 BEP-BEPA	BEP-BEPA : SECONDE
3312022	2/2 BEP-BEPA	BEP-BEPA : TERMINALE
3313011	BACPRO BMA 1 AN	BAC PRO/AGRI, BMA EN 1 AN
3313021	1/2 BACPRO BMA	BAC PRO/AGRI, BMA EN 2 ANS : PREMIERE
3313022	2/2 BACPRO BMA	BAC PRO/AGRI, BMA EN 2 ANS : TERMINALE
3313031	1/3 BACPRO BMA	BAC PRO/AGRI, BMA EN 3 ANS : SECONDE
3313032	2/3 BACPRO BMA	BAC PRO/AGRI, BMA EN 3 ANS : PREMIERE
3313033	3/3 BACPRO BMA	BAC PRO/AGRI, BMA EN 3 ANS : TERMINALE
3313111	BACPRO/A 1 AN	BAC PRO/AGRI EN 1 AN
3313121	1/2 BACPRO/A	BAC PRO/AGRI EN 2 ANS : 1ERE ANNEE
3313122	2/2 BACPRO/A	BAC PRO/AGRI EN 2 ANS : 2EME ANNEE
3313131	1/3 BACPRO/A	BAC PRO/AGRI EN 3 ANS : 1ERE ANNEE
3313132	2/3 BACPRO/A	BAC PRO/AGRI EN 3 ANS : 2EME ANNEE
3313133	3/3 BACPRO/A	BAC PRO/AGRI EN 3 ANS : 3EME ANNEE
3313211	BMA 1 AN	BMA EN 1 AN
3313221	1/2 BMA	BMA EN 2 ANS : 1ERE ANNEE
3313222	2/2 BMA	BMA EN 2 ANS : 2EME ANNEE
3313231	1/3 BMA	BMA EN 3 ANS : 1ERE ANNEE
3313232	2/3 BMA	BMA EN 3 ANS : 2EME ANNEE
3313233	3/3 BMA	BMA EN 3 ANS : 3EME ANNEE
3314011	MC	MENTION COMPLEMENTAIRE
3314111	MC5	MENTION COMPLEMENTAIRE NIVEAU V

3314211	MC4	MENTION COMPLEMENTAIRE NIVEAU IV
3315021	1/2 BP-BPA	BP-BPA : 1ERE ANNEE
3315022	2/2 BP-BPA	BP-BPA : 2EME ANNEE
3315031	1/3 BP-BPA	BP-BPA EN 3 ANS : 1ERE ANNEE
3315032	2/3 BP-BPA	BP-BPA EN 3 ANS : 2EME ANNEE
3315033	3/3 BP-BPA	BP-BPA EN 3 ANS : 3EME ANNEE
3331111	FC POST NIV5	FC DIPLOMANTE POST NIVEAU V
3331211	CSA NIV5	CERTIFICAT DE SPECIALISATION AGRI NIV V
3332111	FC POST NIV4	FC DIPLOMANTE POST NIVEAU IV
3332211	CSA NIV4	CERTIFICAT DE SPECIALISATION AGRI NIV IV
3381111	FCIL NIV-5	MGI - FCIL NIV V
3382111	FCIL NIV-4	MGI - FCIL NIV IV
3401111	TEC-PRO N4 1 AN	DIPLOME TECHNO-PRO DE NIV IV EN 1 AN
3401121	1/2 TEC-PRO N4	DIP TECHNO-PRO DE NIV IV EN 2 ANS: 1E AN
3401122	2/2 TEC-PRO N4	DIP TECHNO-PRO DE NIV IV EN 2 ANS: 2E AN
3401131	1/3 TEC-PRO N4	DIP TECHNO-PRO DE NIV IV EN 3 ANS: 1E AN
3401132	2/3 TEC-PRO N4	DIP TECHNO-PRO DE NIV IV EN 3 ANS: 2E AN
3401133	3/3 TEC-PRO N4	DIP TECHNO-PRO DE NIV IV EN 3 ANS: 3E AN
3411100	PARAMED NIV5	DIPLOME PARAMEDICAL DE NIVEAU V
3411110	PARAM NIV5 1AN	DIPLOME PARAMEDICAL DE NIVEAU V EN 1AN
3411200	SOCIAL NIV5	DIPLOME SOCIAL DE NIVEAU V
3411210	SOCIAL NIV5 1AN	DIPLOME SOCIAL DE NIVEAU V EN 1 AN
3411220	SOCIAL NIV5 2AN	DIPLOME SOCIAL DE NIVEAU V EN 2 ANS
3412100	SOCIAL NIV4	DIPLOME SOCIAL DE NIVEAU IV
3412110	SOCIAL NIV4 1AN	DIPLOME SOCIAL DE NIVEAU IV EN 1 AN
3412120	SOCIAL NIV4 2AN	DIPLOME SOCIAL DE NIVEAU IV EN 2 ANS
3412121	1/2 SOCIAL NIV4	DIPL SOCIAL NIV IV EN 2 ANS : 1ERE ANNEE
3412122	2/2 SOCIAL NIV4	DIPL SOCIAL NIV IV EN 2 ANS : 2EME ANNEE
3412131	1/3 SOCIAL NIV4	DIPL SOCIAL NIV IV EN 3 ANS : 1ERE ANNEE
3412132	2/3 SOCIAL NIV4	DIPL SOCIAL NIV IV EN 3 ANS : 2EME ANNEE
3412133	3/3 SOCIAL NIV4	DIPL SOCIAL NIV IV EN 3 ANS : 3EME ANNEE
3431100	TH NIV5	TITRE HOMOLOGUE DE NIVEAU V
3431200	TH NIV4	TITRE HOMOLOGUE DE NIVEAU IV
4111111	DAEU	DIPLOME ACCES AUX ETUDES UNIVERSITAIRES
4111221	1/2 CAPACITE	CAPACITE EN DROIT : 1ERE ANNEE
4111222	2/2 CAPACITE	CAPACITE EN DROIT : 2EME ANNEE
4112111	DELFI DALF	DIPL ETUDE LANG FRANC, DIPL APPRENT L F
4141111	PREPA CPGE	CLASSE PREPA AUX ETUDES SUPERIEURES
4409100	TEC-PRO POST-2D	DIVERSES FORM TEC-PRO POST-SECONDAIRE
4411120	SOCIAL NIV4 2AN	DIPL SOCIAL POST-SECOND(NIV IV) EN 2 ANS
4411121	1/2 SOCIAL NIV4	DIPL SOCIAL POST-SECOND EN 2 ANS : 1E AN
4411122	2/2 SOCIAL NIV4	DIPL SOCIAL POST-SECOND EN 2 ANS : 2E AN
4411130	SOCIAL NIV4 3AN	DIPL SOCIAL POST-SECOND(NIV IV) EN 3 ANS
4411131	1/3 SOCIAL NIV4	DIPL SOCIAL POST-SECOND EN 3 ANS : 1E AN
4411132	2/3 SOCIAL NIV4	DIPL SOCIAL POST-SECOND EN 3 ANS : 2E AN
4411133	3/3 SOCIAL NIV4	DIPL SOCIAL POST-SECOND EN 3 ANS : 3E AN
4412111	BEES 1ER DEGRE	BREVET D'ETAT EDUCATEUR SPORTIF 1E DEGR
4421111	DIV DIPL NIV IV	DIPL ETAB RECONNU ETAT POST-SECOND N IV
4431111	FC POST NIV 4	FORMATION COMPLEMENTAIRE POST NIVEAU IV
4431211	CSA NIV 4	CERTIFICAT DE SPECIALISATION AGRI NIV IV
4432121	1/2 DU NIV4	DIPL D'UNIV POST-SECONDAIRE : 1ERE ANNEE
4432122	2/2 DU NIV4	DIPL D'UNIV POST-SECONDAIRE : 2EME ANNEE
4433111	CERT COMM NIV4	CERTIF ECOLE COMMERCE POST-SECOND (NIV4)
4434121	1/2 CERT ART N4	CERTIF ECOLE ARTS POST-SECOND : 1E AN

4434122	2/2 CERT ART N4	CERTIF ECOLE ARTS POST-SECOND : 2E AN
4439100	ECOLE NIV4	CERTIFICAT D'ECOLE POST-SECOND (NIV IV)
4439120	ECOLE NIV4	CERTIFICAT D'ECOLE POST-SECOND (NIV IV)
4439121	1/2 DIV CERT N4	DIVERS CERTIF ECOLE POST-SECOND : 1E AN
4439122	2/2 DIV CERT N4	DIVERS CERTIF ECOLE POST-SECOND : 2E AN
4441111	MISE NIV STS	CLASSE DE MISE A NIVEAU STS
4441211	PREPA PARAM	ANNEE PREPA AUX ECOLES PARAMEDICALES
4441311	PREPA ART	ANNEE PREPA AUX ECOLES D'ARTS (YC DANSE
4451111	CONC ADM BAC	PREPA CONC ADM FONCTION PUBLIQUE BAC
4461111	FORM FONCT N4	FORMATION FONTIONNAIRES POST-SECONDAIRE
5111120	DEUG	DEUG
5121121	1/2 DIP P-DAUPH	DIPL UNIVERSITE PARIS-DAUPHINE : 1E AN
5121122	2/2 DIP P-DAUPH	DIPL UNIVERSITE PARIS-DAUPHINE : 2E AN
5141121	1/2 CPGE	CPGE : 1ERE ANNEE
5141122	2/2 CPGE	CPGE : 2EME ANNEE
5141220	PREPA ING	CYCLE PREPA INGENIEUR
5141221	1/2 PREPA ING	CYCLE PREPA INGENIEUR : 1ERE ANNEE
5141222	2/2 PREPA ING	CYCLE PREPA INGENIEUR : 2EME ANNEE
5141320	PREPA COMMERCE	CYCLE PREPA ECOLE DE COMMERCE
5141321	1/2 PREPA COMM	CYCLE PREPA ECOLE COMMERCE : 1ERE ANNEE
5141322	2/2 PREPA COMM	CYCLE PREPA ECOLE COMMERCE : 2EME ANNEE
5142121	1/2 CP MST MSG	CERTIF PREPA MST, MSG : 1ERE ANNEE
5142122	2/2 CP MST MSG	CERTIF PREPA MST, MSG : 2EME ANNEE
5143111	1/1 DNAT-DNSEP	ANNEE COMMUNE DNAT-DNSEP
5401111	TEC-PRO N3 1 AN	DIPLOME TECHNO-PRO DE NIV III EN 1 AN
5401121	1/2 TEC-PRO N3	DIP TECHNO-PRO DE NIV III EN 2 ANS:1E AN
5401122	2/2 TEC-PRO N3	DIP TECHNO-PRO DE NIV III EN 2 ANS:2E AN
5401131	1/3 TEC-PRO N3	DIP TECHNO-PRO DE NIV III EN 3 ANS:1E AN
5401132	2/3 TEC-PRO N3	DIP TECHNO-PRO DE NIV III EN 3 ANS:2E AN
5401133	3/3 TEC-PRO N3	DIP TECHNO-PRO DE NIV III EN 3 ANS:3E AN
5402100	DIV TEC-PRO B+1	DIPL TECHNO-PRO NIV TERM BAC+1
5402200	DIV TEC-PRO B+2	DIPL TECHNO-PRO NIV TERM BAC+2
5403100	COMM N VISE B+1	FORM EC. COMMERCE (DIPL N VISE) BAC+1
5403200	COMM N VISE B+2	FORM EC. COMMERCE (DIPL N VISE) BAC+2
5411011	BTS & ASSIM 1AN	BTS, BTSA, DMA EN 1 AN
5411021	1/2 BTS & ASSIM	BTS, BTSA, DMA EN 2 ANS : 1ERE ANNEE
5411022	2/2 BTS & ASSIM	BTS, BTSA, DMA EN 2 ANS : 2EME ANNEE
5411031	1/3 BTS & ASSIM	BTS, DMA, DTS EN 3 ANS : 1ERE ANNEE
5411032	2/3 BTS & ASSIM	BTS, DMA, DTS EN 3 ANS : 2EME ANNEE
5411033	3/3 BTS & ASSIM	BTS, DMA, DTS EN 3 ANS : 3E A (YC DCESF)
5411111	BTS 1 AN	BTS EN 1 AN
5411121	1/2 BTS	BTS EN 2 ANS : 1ERE ANNEE
5411122	2/2 BTS	BTS EN 2 ANS : 2EME ANNEE
5411131	1/3 BTS	BTS EN 3 ANS : 1ERE ANNEE
5411132	2/3 BTS	BTS EN 3 ANS : 2EME ANNEE
5411133	3/3 BTS	BTS EN 3 ANS : 3EME ANNEE
5411211	BTSA 1 AN	BTSA EN 1 AN
5411221	1/2 BTSA	BTSA EN 2 ANS : 1ERE ANNEE
5411222	2/2 BTSA	BTSA EN 2 ANS : 2EME ANNEE
5411331	1/3 DTS	DIPLOME DE TECHNICIEN SUPERIEUR : 1E AN
5411332	2/3 DTS	DIPLOME DE TECHNICIEN SUPERIEUR : 2E AN
5411333	3/3 DTS	DIPLOME DE TECHNICIEN SUPERIEUR : 3E AN
5411411	DCESF	DIPL CONSEILLER ECO SOCIALE ET FAMILIALE
5411511	DMA 1 AN	DIPLOME METIERS D'ARTS EN 1 AN

5411521	1/2 DMA	DIPLOME METIERS D'ARTS EN 2 ANS : 1E AN
5411522	2/2 DMA	DIPLOME METIERS D'ARTS EN 2 ANS : 2E AN
5411531	1/3 DMA	DIPLOME METIERS D'ARTS EN 3 ANS : 1E AN
5412121	1/2 DUT	DUT : 1ERE ANNEE
5412122	2/2 DUT	DUT : 2EME ANNEE
5413120	PARAM NIV3 2 AN	DIPL PARAMEDICAL DE NIVEAU III EN 2 ANS
5413121	1/2 PARAM NIV3	DIPL PARAM DE NIVEAU III EN 2 ANS: 1E AN
5413122	2/2 PARAM NIV3	DIPL PARAM DE NIVEAU III EN 2 ANS: 2E AN
5413130	PARAM NIV3 3 AN	DIPL PARAMEDICAL DE NIVEAU III EN 3 ANS
5413131	1/3 PARAM NIV3	DIPL PARAM DE NIVEAU III EN 3 ANS: 1E AN
5413132	2/3 PARAM NIV3	DIPL PARAM DE NIVEAU III EN 3 ANS: 2E AN
5413133	3/3 PARAM NIV3	DIPL PARAM DE NIVEAU III EN 3 ANS: 3E AN
5413140	PARAM NIV3 4 AN	DIPL PARAMEDICAL DE NIVEAU III EN 4 ANS
5413141	1/4 PARAM NIV3	DIPL PARAM DE NIVEAU III EN 4 ANS: 1E AN
5413142	2/4 PARAM NIV3	DIPL PARAM DE NIVEAU III EN 4 ANS: 2E AN
5413143	3/4 PARAM NIV3	DIPL PARAM DE NIVEAU III EN 4 ANS: 3E AN
5413144	4/4 PARAM NIV3	DIPL PARAM DE NIVEAU III EN 4 ANS: 4E AN
5413211	SOCIAL NIV3 1AN	DIPLOME SOCIAL DE NIVEAU III EN 1 AN
5413230	SOCIAL NIV3 3AN	DIPLOME SOCIAL DE NIVEAU III EN 3 ANS
5413231	1/3 SOCIAL NIV3	DIPL SOCIAL DE NIV III EN 3 ANS : 1E AN
5413232	2/3 SOCIAL NIV3	DIPL SOCIAL DE NIV III EN 3 ANS : 2E AN
5413233	3/3 SOCIAL NIV3	DIPL SOCIAL DE NIV III EN 3 ANS : 3E AN
5413311	PARAM SP N3 1AN	DIPL PARAMEDICAL SPEC NIV III EN 1 AN
5413320	PARAM SP N3 2AN	DIPL PARAMEDICAL SPEC NIV III EN 2 ANS
5413321	1/2 PARAM SP N3	DIPL PARAM SPEC NIV III EN 2 ANS : 1E AN
5413322	2/2 PARAM SP N3	DIPL PARAM SPEC NIV III EN 2 ANS : 2E AN
5414111	DEUG IUP	DEUG IUP
5415121	1/2 DEUST	DEUST : 1ERE ANNEE
5415122	2/2 DEUST	DEUST : 2EME ANNEE (YC DEUST IUP)
5416111	DNTS	DIPL NAT TECHNOLOGIE SPECIALISEE
5417121	1/2 DNAT AV2006	DIPL NAT ARTS ET TECHN AV 2006 : 1E AN
5417122	2/2 DNAT AV2006	DIPL NAT ARTS ET TECHN AV 2006 : 2E AN
5417221	1/2 PROF DANSE	DIPL PROF DANSE ET MUSIQUE : 1ERE ANNEE
5417222	2/2 PROF DANSE	DIPL PROF DANSE ET MUSIQUE : 2EME ANNEE
5417320	DNAP	DIPLOME NATIONAL ARTS PLASTIQUES
5417421	1/2 DUMI	DIPL UNIV MUSICIEN INTERVENANT : 1E AN
5417422	2/2 DUMI	DIPL UNIV MUSICIEN INTERVENANT : 2E AN
5417520	1ER C ARCHI	DIPLOME 1ER CYCLE ARCHITECTURE
5419111	DPECF	DIPL PREPA ETUDES COMPTABLES FINANCIERES
5419240	CLERC NOTAIRE	DIPLOME DE 1ER CLERC DE NOTAIRE
5419241	1/4 CLERC NOT	DIPLOME DE 1ER CLERC DE NOTAIRE : 1E AN
5419242	2/4 CLERC NOT	DIPLOME DE 1ER CLERC DE NOTAIRE : 2E AN
5419243	3/4 CLERC NOT	DIPLOME DE 1ER CLERC DE NOTAIRE : 3E AN
5419244	4/4 CLERC NOT	DIPLOME DE 1ER CLERC DE NOTAIRE : 4E AN
5419320	1ER C NOTAIRE	DIPLOME FIN 1ER CYCLE ECOLES NOTARIAT
5419400	EXPERT AUTO	DIPLOME EXPERT EN AUTOMOBILE
5429111	1/1 ECOLE NIV3	DIV DIPL ECOLE NIV III EN 1 AN
5429121	1/2 ECOLE NIV3	DIV DIPL ECOLE NIV III EN 2 ANS : 1E AN
5429122	2/2 ECOLE NIV3	DIV DIPL ECOLE NIV III EN 2 ANS : 2E AN
5429200	DPC,DPCE,DPCT	DPC, DPCE, DPCT
5431111	FC POST DUT BTS	FORMATION COMPLEMENTAIRE POST DUT OU BTS
5431211	CSA NIV 3	CERTIFICAT DE SPECIALISATION AGRI NIV II
5432121	1/2 DU NIV3	DIPLOME D'UNIV DE NIVEAU III : 1E ANNEE
5432122	2/2 DU NIV3	DIPLOME D'UNIV DE NIVEAU III : 2E ANNEE

5432220	DEUTEC	DEUTEC
5433131	1/3 CERT COM N3	CERTIF ECOLE COMMERCE NIVEAU III : 1E AN
5433132	2/3 CERT COM N3	CERTIF ECOLE COMMERCE NIVEAU III : 2E AN
5433133	3/3 CERT COM N3	CERTIF ECOLE COMMERCE NIVEAU III : 3E AN
5433211	CERT SPE COM N3	CERTIF SPECIALISATION COMMERCE NIV III
5434131	1/3 CERT ART N3	CERTIF ECOLE D'ARTS NIVEAU III : 1E AN
5434132	2/3 CERT ART N3	CERTIF ECOLE D'ARTS NIVEAU III : 2E AN
5434133	3/3 CERT ART N3	CERTIF ECOLE D'ARTS NIVEAU III : 3E AN
5435100	TH NIV3	TITRE HOMOLOGUE DE NIVEAU III
5439131	1/3 DIV CERT N3	DIVERS CERTIF ECOLE NIVEAU III : 1E AN
5439132	2/3 DIV CERT N3	DIVERS CERTIF ECOLE NIVEAU III : 2E AN
5439133	3/3 DIV CERT N3	DIVERS CERTIF ECOLE NIVEAU III : 3E AN
5439211	DIV CERT SPE N3	DIVERS CERTIF SPECIALISATION NIVEAU III
5451111	CONC ADM B2	PREPA CONC ADM FONCTION PUBLIQUE BAC+2
5461111	FORM FONCT B	FORM FONCTIONNAIRES RECR BAC EN 1AN
5461121	1/2 FONCT N3	FORM FONTIONNAIRES DE NIVEAU III : 1E AN
5461122	2/2 FONCT N3	FORM FONTIONNAIRES DE NIVEAU III : 2E AN
5481111	FCIL NIV-3	MGI - FCIL NIV III
6111111	LICENCE	LICENCE
6112111	MAITRISE	MAITRISE
6113111	MSBM DPRBM	MAITRISE SC BIO-MED, DPRBM
6114131	1/3 LIC (LMD)	LICENCE (LMD) : 1ERE ANNEE
6114132	2/3 LIC (LMD)	LICENCE (LMD) : 2EME ANNEE
6114133	3/3 LIC (LMD)	LICENCE (LMD) : 3EME ANNEE
6115121	1/2 MAST INDIF	MASTER INDIFFERENCIE (LMD) : 1ERE ANNEE
6115122	2/2 MAST INDIF	MASTER INDIFFERENCIE (LMD) : 2EME ANNEE
6115220	MASTER RECH	MASTER RECHERCHE (LMD)
6115221	1/2 MAST RECH	MASTER RECHERCHE (LMD) : 1ERE ANNEE
6115222	2/2 MAST RECH	MASTER RECHERCHE (LMD) : 2EME ANNEE
6116111	MAITRISE (LMD)	MAITRISE (LMD)
6121131	1/3 DIPL IEP	DIPLOME IEP : 1ERE ANNEE
6121132	2/3 DIPL IEP	DIPLOME IEP : 2EME ANNEE
6121133	3/3 DIPL IEP	DIPLOME IEP : 3EME ANNEE
6121151	1/5 DIPL IEP	DIPLOME IEP : 1ERE ANNEE
6121152	2/5 DIPL IEP	DIPLOME IEP : 2EME ANNEE
6121153	3/5 DIPL IEP	DIPLOME IEP : 3EME ANNEE
6121154	4/5 DIPL IEP	DIPLOME IEP : 4EME ANNEE
6121155	5/5 DIPL IEP	DIPLOME IEP : 5EME ANNEE
6122131	1/3 MAGISTERE	MAGISTERE : 1ERE ANNEE
6122132	2/3 MAGISTERE	MAGISTERE : 2EME ANNEE
6122133	3/3 MAGISTERE	MAGISTERE : 3EME ANNEE
6123131	1/3 DIPL. ENS	DIPLOME ENS : 1ERE ANNEE
6123132	2/3 DIPL ENS	DIPLOME ENS : 2EME ANNEE
6123133	3/3 DIPL ENS	DIPLOME ENS : 3EME ANNEE
6124131	1/3 DIPL ETAB	DIPLOME ETAB GRADE LICENCE : 1ERE ANNEE
6124132	2/3 DIPL ETAB	DIPLOME ETAB GRADE LICENCE : 2EME ANNEE
6124133	3/3 DIPL ETAB	DIPLOME ETAB GRADE LICENCE : 3EME ANNEE
6125121	1/2 MAST ETAB	MASTER ETABLISSEMENT : 1ERE ANNEE
6125122	2/2 MAST ETAB	MASTER ETABLISSEMENT : 2EME ANNEE
6131141	1/4 EPHE	DIPLOME DE L'EPHE : 1ERE ANNEE
6131142	2/4 EPHE	DIPLOME DE L'EPHE : 2EME ANNEE
6131143	3/4 EPHE	DIPLOME DE L'EPHE : 3EME ANNEE
6131144	4/4 EPHE	DIPLOME DE L'EPHE : 4EME ANNEE
6132131	1/3 EHESS	DIPLOME DE L'EHESS : 1ERE ANNEE

6132132	2/3 EHESS	DIPLOME DE L'EHESS : 2EME ANNEE
6132133	3/3 EHESS	DIPLOME DE L'EHESS : 3EME ANNEE
6139151	1/5 CERT GENE	1/5 DIVERSES CERTIFICATIONS GENERALES
6139152	2/5 CERT GENE	2/5 DIVERSES CERTIFICATIONS GENERALES
6139153	3/5 CERT GENE	3/5 DIVERSES CERTIFICATIONS GENERALES
6139154	4/5 CERT GENE	4/5 DIVERSES CERTIFICATIONS GENERALES
6139155	5/5 CERT GENE	5/5 DIVERSES CERTIFICATIONS GENERALES
6401100	SPEC HORS ING	DIPL ECOLE DE SPECIALISATION (HORS ING)
6402111	TEC-PRO N1-2:1A	DIPLOME TECHNO-PRO NIV I-II:1ERE ANNEE
6402211	TEC-PRO N1-2:2A	DIPLOME TECHNO-PRO NIV I-II:2EME ANNEE
6402311	TEC-PRO N1-2:3A	DIPLOME TECHNO-PRO NIV I-II:3EME ANNEE
6402411	TEC-PRO BAC+6	DIPLOME TECHNO-PRO DE NIVEAU I-II:BAC+6
6403120	CAPE (YC PREPA)	CAPE-1ER DEGRE PRIVE : PREPA & FORMATION
6404100	SPEC ARCHI	SPECIALISATION ECOLES D'ARCHITECTURE
6405100	DIV TEC-PRO B+3	DIPL TECHNO-PRO NIV TERM BAC+3
6405200	DIV TEC-PRO B+4	DIPL TECHNO-PRO NIV TERM BAC+4
6405300	DIV TEC-PRO B+5	DIPL TECHNO-PRO NIV TERM BAC+5
6405400	DIV TEC-PRO B+6	DIPL TECHNO-PRO NIV TERM BAC+6
6406100	COMM N VISE B+3	FORM EC. COMMERCE (DIPL N VISE) BAC+3
6406200	COMM N VISE B+4	FORM EC. COMMERCE (DIPL N VISE) BAC+4
6406300	COMM N VISE B+5	FORM EC. COMMERCE (DIPL N VISE) BAC+5
6406400	COMM N VISE B+6	FORM EC. COMMERCE (DIPL N VISE) BAC+6
6411111	LICENCE IUP	LICENCE IUP
6411211	LICENCE PRO	LICENCE PROFESSIONNELLE
6412111	MAITRISE IUP	MAITRISE IUP
6412221	1/2 MST	MST : 1ERE ANNEE
6412222	2/2 MST	MST : 2EME ANNEE
6412321	1/2 MSG	MSG : 1ERE ANNEE
6412322	2/2 MSG	MSG : 2EME ANNEE
6412421	1/2 MIAGE	MIAGE : 1ERE ANNEE
6412422	2/2 MIAGE	MIAGE : 2EME ANNEE
6412511	ING MAITRE	TITRE D'INGENIEUR-MAITRE
6413111	DESS	DESS
6413221	1/2 DRT	DRT : 1ERE ANNEE
6413222	2/2 DRT	DRT : 2EME ANNEE
6414161	1E AN MED DENT	CURSUS DE MEDECINE - ODONTO : 1ERE ANNEE
6414162	2/6 MEDECINE	CURSUS DE MEDECINE : 2EME ANNEE
6414163	3/6 MEDECINE	CURSUS DE MEDECINE : 3EME ANNEE
6414164	4/6 MEDECINE	CURSUS DE MEDECINE : 4EME ANNEE
6414165	5/6 MEDECINE	CURSUS DE MEDECINE : 5EME ANNEE
6414166	6/6 MEDECINE	CURSUS DE MEDECINE : 6EME ANNEE
6414177	QUAL MED GENE	CURSUS DE MEDECINE : QUAL MED GENE
6414188	DES MEDECINE	CURSUS DE MEDECINE : DES
6414199	THESE MEDECINE	CURSUS DE MEDECINE : THESE
6414252	2/5 ODONTO	CURSUS D'ODONTOLOGIE : 2EME ANNEE
6414253	3/5 ODONTO	CURSUS D'ODONTOLOGIE : 3EME ANNEE
6414254	4/5 ODONTO	CURSUS D'ODONTOLOGIE : 4EME ANNEE
6414255	5/5 ODONTO	CURSUS D'ODONTOLOGIE : 5EME ANNEE
6414266	DE CH DENT	CURSUS D'ODONTOLOGIE : DE CHIRURGIE DEN
6414277	CESMO	CURSUS D'ODONTOLOGIE : CESMO
6414299	THESE ODONTO	CURSUS D'ODONTOLOGIE : THESE
6414351	1/5 PHARMA	CURSUS DE PHARMACIE : 1ERE ANNEE
6414352	2/5 PHARMA	CURSUS DE PHARMACIE : 2EME ANNEE
6414353	3/5 PHARMA	CURSUS DE PHARMACIE : 3EME ANNEE

6414354	4/5 PHARMA	CURSUS DE PHARMACIE : 4EME ANNEE
6414355	5/5 PHARMA	CURSUS DE PHARMACIE : 5EME ANNEE
6414366	DE PHARMA	CURSUS DE PHARMACIE : DE DOCTEUR PHARMA
6414377	DES PHARMA	CURSUS DE PHARMACIE : DES
6414399	THESE PHARMA	CURSUS DE PHARMACIE : THESE
6414400	DESC CAPA MED	DIPL DE SANTE COMPL : DESC, CAPA MEDECIN
6414500	DIPL MED ETRANG	DIPL DE SANTE COMPL POUR MEDECINS ETRANG
6414600	AEA CES DESCB	DIPL DE SANTE COMPL : AEA CES DENT DESCB
6414761	PACES	CURSUS DE SANTE: 1ERE AN COMMUNE
6414840	SAGE-FEMME	DIPL SAGE-FEMME EN 4 ANS
6414841	1/4 SAGE-FEMME	DIPL SAGE-FEMME EN 4 ANS:1E AN
6414842	2/4 SAGE-FEMME	DIPL SAGE-FEMME EN 4 ANS:2E AN
6414843	3/4 SAGE-FEMME	DIPL SAGE-FEMME EN 4 ANS:3E AN
6414844	4/4 SAGE-FEMME	DIPL SAGE-FEMME EN 4 ANS:4E AN
6415150	DOCT VETO	DIPLOME DE DOCTEUR VETERINAIRE
6415151	1/5 VETO	DIPLOME DE DOCTEUR VETERINAIRE : 1E AN
6415152	2/5 VETO	DIPLOME DE DOCTEUR VETERINAIRE : 2E AN
6415153	3/5 VETO	DIPLOME DE DOCTEUR VETERINAIRE : 3E AN
6415154	4/5 VETO	DIPLOME DE DOCTEUR VETERINAIRE : 4E AN
6415155	5/5 VETO	DIPLOME DE DOCTEUR VETERINAIRE : 5E AN
6415230	DES VETO	DIPLOME ETUDES SPECIALISEES VETERINAIRE
6415340	DEF VETO	DIPLOME ETUDES FONDAMENTALES VETERINAIRE
6416121	1/2 ARCHI NIV 1	DIPLOME D'ARCHITECTE: 1E AN MASTER
6416122	2/2 ARCHI NIV 1	DIPLOME D'ARCHITECTE: 2E AN MASTER
6416131	1/3 ARCHI NIV 2	DIPLOME D'ETUDES EN ARCHITECTURE: 1E AN
6416132	2/3 ARCHI NIV 2	DIPLOME D'ETUDES EN ARCHITECTURE: 2E AN
6416133	3/3 ARCHI NIV 2	DIPLOME D'ETUDES EN ARCHITECTURE: 3E AN
6416160	ARCHI	DIPLOME D'ARCHITECTE
6416161	1/6 ARCHI	DIPLOME D'ARCHITECTE : 1ERE ANNEE
6416162	2/6 ARCHI	DIPLOME D'ARCHITECTE : 2EME ANNEE
6416163	3/6 ARCHI	DIPLOME D'ARCHITECTE : 3EME ANNEE
6416164	4/6 ARCHI	DIPLOME D'ARCHITECTE : 4EME ANNEE
6416165	5/6 ARCHI	DIPLOME D'ARCHITECTE : 5EME ANNEE
6416166	6/6 ARCHI	DIPLOME D'ARCHITECTE : 6EME ANNEE
6416211	1/1 DNMADE	DIPL NAT MET ART DESIGN EN 1 AN
6416221	1/2 DSAA	DIPL SUP ARTS APPLIQUES : 1E AN
6416222	2/2 DSAA	DIPL SUP ARTS APPLIQUES : 2E AN
6416231	1/3 DNMADE	DIPL NAT MET ART DESIGN EN 3 ANS : 1E AN
6416232	2/3 DNMADE	DIPL NAT MET ART DESIGN EN 3 ANS : 2E AN
6416233	3/3 DNMADE	DIPL NAT MET ART DESIGN EN 3 ANS : 3E AN
6416341	1/4 DNSEP	DIPL NAT SUP EXPRESSION PLASTIQUE: 1E AN
6416342	2/4 DNSEP	DIPL NAT SUP EXPRESSION PLASTIQUE: 2E AN
6416343	3/4 DNSEP	DIPL NAT SUP EXPRESSION PLASTIQUE: 3E AN
6416344	4/4 DNSEP	DIPL NAT SUP EXPRESSION PLASTIQUE: 4E AN
6416440	PAYSAGISTE	DIPLOME DE PAYSAGISTE
6416441	1/4 PAYSAGISTE	DIPLOME DE PAYSAGISTE : 1ERE ANNEE
6416442	2/4 PAYSAGISTE	DIPLOME DE PAYSAGISTE : 2EME ANNEE
6416443	3/4 PAYSAGISTE	DIPLOME DE PAYSAGISTE : 3EME ANNEE
6416444	4/4 PAYSAGISTE	DIPLOME DE PAYSAGISTE : 4EME ANNEE
6416520	2E C ARCHI	DIPLOME 2EME CYCLE ARCHITECTURE
6416600	CAFP EC MUSIQUE	CERT APT FONC PROF ECOLES MUSIQUE AGREES
6416730	3E AN ARCHI	DIPLOME D'ETUDES EN ARCHITECTURE
6416821	1/2 DNAT	DIPL NATIONAL ARTS ET TECHNIQUES : 1E AN
6416822	2/2 DNAT	DIPL NATIONAL ARTS ET TECHNIQUES : 2E AN

6416900	DIPL ARCHI	DIPLOMES D'ARCHITECTURE
6417120	MASTER PRO	MASTER PROFESSIONNEL (LMD)
6417121	1/2 MAST PRO	MASTER PROFESSIONNEL (LMD) : 1ERE ANNEE
6417122	2/2 MAST PRO	MASTER PROFESSIONNEL (LMD) : 2EME ANNEE
6417221	1/2 MAST ING	MASTER INGENIEUR (LMD) : 1ERE ANNEE
6417222	2/2 MAST ING	MASTER INGENIEUR (LMD) : 2EME ANNEE
6417321	1/2 MAST ENS	MASTER ENSEIGNEMENT(LMD):1ERE ANNEE
6417322	2/2 MAST ENS	MASTER ENSEIGNEMENT(LMD):2EME ANNEE
6418130	PARAM NIV 2	DIPL PARAM DE NIVEAU II EN 3 ANS
6418131	1/3 PARAM NIV 2	DIPL PARAM DE NIVEAU II EN 3 ANS : 1E AN
6418132	2/3 PARAM NIV 2	DIPL PARAM DE NIVEAU II EN 3 ANS : 2E AN
6418133	3/3 PARAM NIV 2	DIPL PARAM DE NIVEAU II EN 3 ANS : 3E AN
6418220	PARAM SP N1	DIPL PARAM SPEC NIV I EN 2 ANS
6418221	1/2 PARAM SP N1	DIPL PARAM SPEC NIV I EN 2 ANS : 1E AN
6418222	2/2 PARAM SP N1	DIPL PARAM SPEC NIV I EN 2 ANS : 2E AN
6419120	DECF	DIPL ETUDES COMPTABLES FINANCIERES
6419131	1/3 DCG	DIPLOME COMPTABILITE ET GESTION : 1E AN
6419132	2/3 DCG	DIPLOME COMPTABILITE ET GESTION : 2E AN
6419133	3/3 DCG	DIPLOME COMPTABILITE ET GESTION : 3E AN
6419211	DESCF	DIPL ETUDES SUP COMPTABLES FINANCIERES
6419221	1/2 DSCG	DIPL SUP COMPTABILITE ET GESTION : 1E AN
6419222	2/2 DSCG	DIPL SUP COMPTABILITE ET GESTION : 2E AN
6419311	DS NOTARIAT	DIPLOME SUPERIEUR DE NOTARIAT
6419421	1/2 CEJ	CERTIFICAT ETUDES JUDICIAIRES : 1E ANNEE
6419422	2/2 CEJ	CERTIFICAT ETUDES JUDICIAIRES : 2E ANNEE
6419511	GUIDE INTERP	DIPL NAT GUIDE INTERPRETE NATIONAL
6419621	1/2 OENOLOGIE	DIPLOME NATIONAL D'OENOLOGIE : 1E ANNEE
6419622	2/2 OENOLOGIE	DIPLOME NATIONAL D'OENOLOGIE : 2E ANNEE
6419750	NAVIG MARITIME	BREVET DE CAPITAINE NAVIGATION MARITIME
6419751	1/5 NAVIG MARIT	BREVET CAPITAINE NAVIG MARITIME : 1E AN
6419752	2/5 NAVIG MARIT	BREVET CAPITAINE NAVIG MARITIME : 2E AN
6419753	3/5 NAVIG MARIT	BREVET CAPITAINE NAVIG MARITIME : 3E AN
6419754	4/5 NAVIG MARIT	BREVET CAPITAINE NAVIG MARITIME : 4E AN
6419755	5/5 NAVIG MARIT	BREVET CAPITAINE NAVIG MARITIME : 5E AN
6419811	SOCIAL NIV2	DIPLOME SOCIAL DE NIVEAU II
6419831	1/3 SOCIAL NIV2	DIPLOME SOCIAL DE NIVEAU II :1E ANNEE
6419832	2/3 SOCIAL NIV2	DIPLOME SOCIAL DE NIVEAU II :2E ANNEE
6419833	3/3 SOCIAL NIV2	DIPLOME SOCIAL DE NIVEAU II :2E ANNEE
6419921	1/2 CA SOURDS	CERTIF APTITUDE ENSGT JEUNES SOURDS:1E A
6419922	2/2 CA SOURDS	CERTIF APTITUDE ENSGT JEUNES SOURDS:2E A
6421031	1/3 ING YC FIP	DIPL INGENIEUR EN 3 ANS (yc FIP) :1E AN
6421032	2/3 ING YC FIP	DIPL INGENIEUR EN 3 ANS (yc FIP) :2E AN
6421033	3/3 ING YC FIP	DIPL INGENIEUR EN 3 ANS (yc FIP) :3E AN
6421130	ING 3 ANS	DIPLOME D'INGENIEUR CLASSIQUE EN 3 ANS
6421131	1/3 ING	DIPL INGENIEUR CLASSIQUE EN 3 ANS: 1E AN
6421132	2/3 ING	DIPL INGENIEUR CLASSIQUE EN 3 ANS: 2E AN
6421133	3/3 ING	DIPL INGENIEUR CLASSIQUE EN 3 ANS: 3E AN
6421140	ING 4 ANS	DIPLOME D'INGENIEUR CLASSIQUE EN 4 ANS
6421141	1/4 ING	DIPL INGENIEUR CLASSIQUE EN 4 ANS: 1E AN
6421142	2/4 ING	DIPL INGENIEUR CLASSIQUE EN 4 ANS: 2E AN
6421143	3/4 ING	DIPL INGENIEUR CLASSIQUE EN 4 ANS: 3E AN
6421144	4/4 ING	DIPL INGENIEUR CLASSIQUE EN 4 ANS: 4E AN
6421150	ING 5 ANS	DIPLOME D'INGENIEUR CLASSIQUE EN 5 ANS
6421151	1/5 ING	DIPL INGENIEUR CLASSIQUE EN 5 ANS: 1E AN

6421152	2/5 ING	DIPL INGENIEUR CLASSIQUE EN 5 ANS: 2E AN
6421153	3/5 ING	DIPL INGENIEUR CLASSIQUE EN 5 ANS: 3E AN
6421154	4/5 ING	DIPL INGENIEUR CLASSIQUE EN 5 ANS: 4E AN
6421155	5/5 ING	DIPL INGENIEUR CLASSIQUE EN 5 ANS: 5E AN
6421200	FIP (EX NFI) FC	DIPL ING PARTENARIAT (EX NFI) FORM CONT
6421230	FIP(EX NFI) 3AN	DIPL ING PARTENARIAT (EX NFI) EN 3 ANS
6421231	1/3 FIP(EX NFI)	DIPL ING PARTEN (EX NFI) EN 3 ANS: 1E AN
6421232	2/3 FIP(EX NFI)	DIPL ING PARTEN (EX NFI) EN 3 ANS: 2E AN
6421233	3/3 FIP(EX NFI)	DIPL ING PARTEN (EX NFI) EN 3 ANS: 3E AN
6421250	FIP(EX NFI) 5AN	DIPL ING PARTENARIAT (EX NFI) EN 5 ANS
6421251	1/5 FIP(EX NFI)	DIPL ING PARTEN (EX NFI) EN 5 ANS: 1E AN
6421252	2/5 FIP(EX NFI)	DIPL ING PARTEN (EX NFI) EN 5 ANS: 2E AN
6421253	3/5 FIP(EX NFI)	DIPL ING PARTEN (EX NFI) EN 5 ANS: 3E AN
6421254	4/5 FIP(EX NFI)	DIPL ING PARTEN (EX NFI) EN 5 ANS: 4E AN
6421255	5/5 FIP(EX NFI)	DIPL ING PARTEN (EX NFI) EN 5 ANS: 5E AN
6421311	ING SPEC 1 AN	DIPL INGENIEUR SPECIALISATION EN 1 AN
6421320	ING SPEC 2 ANS	DIPL INGENIEUR SPECIALISATION EN 2 ANS
6421321	1/2 ING SPEC	DIPL INGENIEUR SPECIALIS EN 2 ANS: 1E AN
6421322	2/2 ING SPEC	DIPL INGENIEUR SPECIALIS EN 2 ANS: 2E AN
6421400	ING POST-DEST	INGENIEUR POST-DEST (DECRET DU 04/01/64)
6421500	ING POST-DUT	INGENIEUR POST-DUT (ARRETE DU 31/07/74)
6421600	ING DIPLOM ETAT	INGENIEUR DIPLOME PAR ETAT (DECRET 1975)
6421900	AUTRES DIPLOMES	AUTRES DIPLOMES
6422000	COM VISE NC ESC	DIPL ECOLE DE COMMERCE VISE (NC ESC ICN)
6422100	COM RECR B2	DIPLOME COMMERCE VISE RECRUT PREPA
6422130	ESC ICN	DIPLOME DES ESC ET DE L'ICN
6422131	1/3 COM RECR B2	1/3 DIPLOME COMMERCE VISE RECRUT PREPA
6422132	2/3 COM RECR B2	2/3 DIPLOME COMMERCE VISE RECRUT PREPA
6422133	3/3 COM RECR B2	3/3 DIPLOME COMMERCE VISE RECRUT PREPA
6422200	COM RECR BAC	DIPLOME COMMERCE VISE RECRUTANT BAC
6422230	COM RECR B 3AN	DIPL COMMERCE VISE RECRUT BAC EN 3 ANS
6422231	1/3 COM RECR B	1/3 DIPLOME COMMERCE VISE RECRUTANT BAC
6422232	2/3 COM RECR B	2/3 DIPLOME COMMERCE VISE RECRUTANT BAC
6422233	3/3 COM RECR B	3/3 DIPLOME COMMERCE VISE RECRUTANT BAC
6422250	COM RECR B 5AN	DIPL COMMERCE VISE RECRUT BAC EN 5 ANS
6422251	1/5 COM RECR B	1/5 DIPLOME COMMERCE VISE RECRUTANT BAC
6422252	2/5 COM RECR B	2/5 DIPLOME COMMERCE VISE RECRUTANT BAC
6422253	3/5 COM RECR B	3/5 DIPLOME COMMERCE VISE RECRUTANT BAC
6422254	4/5 COM RECR B	4/5 DIPLOME COMMERCE VISE RECRUTANT BAC
6422255	5/5 COM RECR B	5/5 DIPLOME COMMERCE VISE RECRUTANT BAC
6422300	DIPL SPEC COMM	DIPLOME SPECIALISATION COMMERCE VISE
6423130	ART RECR B2	DIPLOME ECOLE D'ARTS RECRUTANT BAC+2
6423131	1/3 ART RECR B2	1/3 DIPLOME ECOLE D'ARTS RECRUTANT BAC+
6423132	2/3 ART RECR B2	2/3 DIPLOME ECOLE D'ARTS RECRUTANT BAC+
6423133	3/3 ART RECR B2	3/3 DIPLOME ECOLE D'ARTS RECRUTANT BAC+
6423250	ART RECR BAC	DIPLOME ECOLE D'ARTS RECRUTANT BAC
6423251	1/5 ART RECR B	1/5 DIPLOME ECOLE D'ARTS RECRUTANT BAC
6423252	2/5 ART RECR B	2/5 DIPLOME ECOLE D'ARTS RECRUTANT BAC
6423253	3/5 ART RECR B	3/5 DIPLOME ECOLE D'ARTS RECRUTANT BAC
6423254	4/5 ART RECR B	4/5 DIPLOME ECOLE D'ARTS RECRUTANT BAC
6423255	5/5 ART RECR B	5/5 DIPLOME ECOLE D'ARTS RECRUTANT BAC
6429151	1/5 ENS BLOIS	DIPL ENS NATURE PAYSAGE DE BLOIS : 1E AN
6429152	2/5 ENS BLOIS	DIPL ENS NATURE PAYSAGE DE BLOIS : 2E AN
6429153	3/5 ENS BLOIS	DIPL ENS NATURE PAYSAGE DE BLOIS : 3E AN

6429154	4/5 ENS BLOIS	DIPL ENS NATURE PAYSAGE DE BLOIS : 4E AN
6429155	5/5 ENS BLOIS	DIPL ENS NATURE PAYSAGE DE BLOIS : 5E AN
6429230	DIV DIPL B2	DIVERS DIPLOMES ECOLE RECRUT BAC+2
6429231	1/3 DIV DIPL B2	1/3 DIVERS DIPLOMES ECOLE RECRUT BAC+2
6429232	2/3 DIV DIPL B2	2/3 DIVERS DIPLOMES ECOLE RECRUT BAC+2
6429233	3/3 DIV DIPL B2	3/3 DIVERS DIPLOMES ECOLE RECRUT BAC+2
6429351	1/5 DIV DIPL B	1/5 DIVERS DIPLOMES ECOLE RECRUTANT BAC
6429352	2/5 DIV DIPL B	2/5 DIVERS DIPLOMES ECOLE RECRUTANT BAC
6429353	3/5 DIV DIPL B	3/5 DIVERS DIPLOMES ECOLE RECRUTANT BAC
6429354	4/5 DIV DIPL B	4/5 DIVERS DIPLOMES ECOLE RECRUTANT BAC
6429355	5/5 DIV DIPL B	5/5 DIVERS DIPLOMES ECOLE RECRUTANT BAC
6429411	DIV DIPL SPEC	DIVERS DIPLOMES SPECIALISATION NIV I-II
6429500	DEST,DESE,DESA	DEST, DESE, DESA
6429600	DIP C DU CNAM	DIPLOME C DU CNAM
6429711	DIV DIPL B5 1AN	DIVERS DIPLOMES ECOLE NIV BAC+5 EN 1 AN
6429721	1/2 DIV DIPL B5	1/2 DIVERS DIPLOMES ECOLE NIV BAC+5
6429722	2/2 DIV DIPL B5	2/2 DIVERS DIPLOMES ECOLE NIV BAC+5
6429731	1/3 DIV DIPL B5	1/3 DIVERS DIPLOMES ECOLE NIV BAC+5
6429732	2/3 DIV DIPL B5	2/3 DIVERS DIPLOMES ECOLE NIV BAC+5
6429733	3/3 DIV DIPL B5	3/3 DIVERS DIPLOMES ECOLE NIV BAC+5
6429811	DIV DIPL B4	DIVERS DIPLOMES ECOLE BAC+4
6429821	1/2 DIPL BAC+4	1/2 DIVERS DIPLOMES ECOLE NIV BAC+4
6429822	2/2 DIPL BAC+4	2/2 DIVERS DIPLOMES ECOLE NIV BAC+4
6429831	1/3 DIPL BAC+4	1/3 DIVERS DIPLOMES ECOLE NIV BAC+4
6429832	2/3 DIPL BAC+4	2/3 DIVERS DIPLOMES ECOLE NIV BAC+4
6429833	3/3 DIPL BAC+4	3/3 DIVERS DIPLOMES ECOLE NIV BAC+4
6429841	1/4 DIPL BAC+4	1/4 DIVERS DIPLOMES ECOLE NIV BAC+4
6429842	2/4 DIPL BAC+4	2/4 DIVERS DIPLOMES ECOLE NIV BAC+4
6429843	3/4 DIPL BAC+4	3/4 DIVERS DIPLOMES ECOLE NIV BAC+4
6429844	4/4 DIPL BAC+4	4/4 DIVERS DIPLOMES ECOLE NIV BAC+4
6429911	DIV DIPL B3	DIVERS DIPLOMES ECOLE BAC+3
6429931	1/3 DIPL BAC+3	1/3 DIVERS DIPLOMES ECOLE NIV BAC+3
6429932	2/3 DIPL BAC+3	2/3 DIVERS DIPLOMES ECOLE NIV BAC+3
6429933	3/3 DIPL BAC+3	3/3 DIVERS DIPLOMES ECOLE NIV BAC+3
6432011	MASTERE MBA	MASTERE, MBA
6433121	1/2 DU NIV2	DIPLOME D'UNIV DE NIVEAU II : 1ERE ANNE
6433122	2/2 DU NIV2	DIPLOME D'UNIV DE NIVEAU II : 2EME ANNE
6433221	1/2 DU NIV1	DIPLOME D'UNIV DE NIVEAU I : 1ERE ANNEE
6433222	2/2 DU NIV1	DIPLOME D'UNIV DE NIVEAU I : 2EME ANNEE
6434131	1/3 CERT COM B2	1/3 CERTIF ECOLE COMMERCE RECRUT BAC+2
6434132	2/3 CERT COM B2	2/3 CERTIF ECOLE COMMERCE RECRUT BAC+2
6434133	3/3 CERT COM B2	3/3 CERTIF ECOLE COMMERCE RECRUT BAC+2
6434200	CERT COM REC B	CERTIF ECOLE COMMERCE RECRUT BAC
6434240	CERT COM RECR B	CERTIF ECOLE COMMERCE RECRUTANT BAC
6434241	1/4 CERT COM B	1/4 CERTIF ECOLE COMMERCE RECRUTANT BAC
6434242	2/4 CERT COM B	2/4 CERTIF ECOLE COMMERCE RECRUTANT BAC
6434243	3/4 CERT COM B	3/4 CERTIF ECOLE COMMERCE RECRUTANT BAC
6434244	4/4 CERT COM B	4/4 CERTIF ECOLE COMMERCE RECRUTANT BAC
6434251	1/5 CERT COM B	1/5 CERTIF ECOLE COMMERCE RECRUTANT BAC
6434252	2/5 CERT COM B	2/5 CERTIF ECOLE COMMERCE RECRUTANT BAC
6434253	3/5 CERT COM B	3/5 CERTIF ECOLE COMMERCE RECRUTANT BAC
6434254	4/5 CERT COM B	4/5 CERTIF ECOLE COMMERCE RECRUTANT BAC
6434255	5/5 CERT COM B	5/5 CERTIF ECOLE COMMERCE RECRUTANT BAC
6434321	1/2 SPE COM N12	1/2 CERTIF SPECIALIS COMMERCE NIV I-II

6434322	2/2 SPE COM N12	2/2 CERTIF SPECIALIS COMMERCE NIV I-II
6435121	1/2 CERT ART B2	1/2 CERTIF ECOLE D'ARTS RECRUTANT BAC+2
6435122	2/2 CERT ART B2	2/2 CERTIF ECOLE D'ARTS RECRUTANT BAC+2
6435250	CERT ART RECR B	CERTIF ECOLE D'ARTS RECRUTANT BAC
6435251	1/5 CERT ART B	1/5 CERTIF ECOLE D'ARTS RECRUTANT BAC
6435252	2/5 CERT ART B	2/5 CERTIF ECOLE D'ARTS RECRUTANT BAC
6435253	3/5 CERT ART B	3/5 CERTIF ECOLE D'ARTS RECRUTANT BAC
6435254	4/5 CERT ART B	4/5 CERTIF ECOLE D'ARTS RECRUTANT BAC
6435255	5/5 CERT ART B	5/5 CERTIF ECOLE D'ARTS RECRUTANT BAC
6435311	CERT SPEC ART	CERTIFICAT SPECIALISATION ECOLE D'ARTS
6435400	CERT ART NIV 2	CERTIFICAT ECOLE D'ARTS DE NIVEAU II
6435500	CERT ART NIV 1	CERTIFICAT ECOLE D'ARTS DE NIVEAU I
6436160	DIPL OSTEOPATHE	DIPLOME D'OSTEOPATHE
6436161	1/6 OSTEOPATHE	DIPLOME D'OSTEOPATHE : 1ERE ANNEE
6436162	2/6 OSTEOPATHE	DIPLOME D'OSTEOPATHE : 2EME ANNEE
6436163	3/6 OSTEOPATHE	DIPLOME D'OSTEOPATHE : 3EME ANNEE
6436164	4/6 OSTEOPATHE	DIPLOME D'OSTEOPATHE : 4EME ANNEE
6436165	5/6 OSTEOPATHE	DIPLOME D'OSTEOPATHE : 5EME ANNEE
6436166	6/6 OSTEOPATHE	DIPLOME D'OSTEOPATHE : 6EME ANNEE
6437100	TH NIV2	TITRE HOMOLOGUE DE NIVEAU II
6437200	TH NIV1	TITRE HOMOLOGUE DE NIVEAU I
6439000	ECOLE TECHN SUP	CERTIFICAT ECOLE TECHN SUP (ING NON HAB)
6439100	ECOLE TECHN SUP	CERTIFICAT ECOLE TECHN SUP (ING NON HAB)
6439130	DIV CERT REC B2	DIVERS CERTIFICAT ECOLE RECRUTANT BAC+2
6439131	1/3 DIV CERT B2	1/3 DIVERS CERTIF ECOLE RECRUTANT BAC+2
6439132	2/3 DIV CERT B2	2/3 DIVERS CERTIF ECOLE RECRUTANT BAC+2
6439133	3/3 DIV CERT B2	3/3 DIVERS CERTIF ECOLE RECRUTANT BAC+2
6439250	DIV CERT REC B	DIVERS CERTIFICAT ECOLE RECRUTANT BAC
6439251	1/5 DIV CERT B	1/5 DIVERS CERTIF ECOLE RECRUTANT BAC
6439252	2/5 DIV CERT B	2/5 DIVERS CERTIF ECOLE RECRUTANT BAC
6439253	3/5 DIV CERT B	3/5 DIVERS CERTIF ECOLE RECRUTANT BAC
6439254	4/5 DIV CERT B	4/5 DIVERS CERTIF ECOLE RECRUTANT BAC
6439255	5/5 DIV CERT B	5/5 DIVERS CERTIF ECOLE RECRUTANT BAC
6439321	1/2 DIV SPE N2	1/2 DIVERS CERTIF SPECIALIS NIV I-II
6439322	2/2 DIV SPE N2	2/2 DIVERS CERTIF SPECIALIS NIV I-II
6441111	PREP RECRUT B4	PREPA ENTREE FORM SELECTIVE RECRUT BAC+4
6451111	PREPA CAPE PUBL	PREPA CAPE - 1ER DEGRE PUBLIC
6451211	PREPA CAPE PRIV	PREPA CAPE - 1ER DEGRE PRIVE
6452111	PREPA CAPES	PREPA CAPES, CAPEPS
6452211	PREPA CAPET	PREPA CAPET
6452311	PREPA CAPLP	PREPA CAPLP
6452411	PREPA CPE	PREPA CPE (CONSEILLER PRINCIPAL EDUC)
6452511	PREPA CAFEP	PREPA CAFEP (ENSEIGNANT 2D DEGRE PRIVE)
6452611	PREPA AGREG	PREPA AGREGATION
6452711	PREPA CON ENSEI	PREPA CONCOURS ENSEIGNEMENT
6453111	CONC ADM B3	PREPA CONC ADM FONCTION PUBLIQUE BAC+3
6453211	CONC ADM B4	PREPA CONC ADM FONCTION PUBLIQUE BAC+4
6454121	1/2 CER ENS SP	CERTIF APT ENSEIGNT SPEC 1ERE ANNEE
6454122	2/2 CER ENS SP	CERTIF APT ENSEIGNT SPEC 2EME ANNEE
6461111	CAPE PUBLIC	FORMATION ENSEIGNANTS 1ER DEGRE PUBLIC
6461211	SPEC ENST 1D	SPEC ENST 1ER DEGRE (CAPSAIS, CAPA-SH)
6461311	PSY SCO	DIPLOME D' ETAT DE PSYCHOLOGUE SCOLAIR
6461411	CAPE PRIVE	FORMATION ENSEIGNANTS 1ER DEGRE PRIVE
6462111	CAPES CAPEPS	CAPES, CAPEPS

6462211	CAPET	CAPET
6462311	CAPLP	CAPLP
6462411	CPE	CPE (CONSEILLER PRINCIPAL EDUC)
6462511	CAFEP	CAFEP (ENSEIGNANT 2D DEGRE PRIVE)
6462611	AGREGATION	AGREGATION
6462711	ENST 2D AGRI	FORMATION ENSEIGNANTS 2D DEGRE AGRICOLE
6462811	SPEC ENST 2D	SPEC ENST 2D DEGRE (2CA-SH)
6463131	1/3 FONCT B2	1/3 F.FONCTIONNAIRE RECR BAC+2 (NC ENST)
6463132	2/3 FONCT B2	2/3 F.FONCTIONNAIRE RECR BAC+2 (NC ENST)
6463133	3/3 FONCT B2	3/3 F.FONCTIONNAIRE RECR BAC+2 (NC ENST)
6463221	1/2 FONCT B3	1/2 F.FONCTIONNAIRE RECR BAC+3 (NC ENST)
6463222	2/2 FONCT B3	2/2 F.FONCTIONNAIRE RECR BAC+3 (NC ENST)
6463300	F. FONCT B4	F. FONCTIONNAIRE RECR BAC+4 (NC ENST)
6463331	1/3 FONCT B4	1/3 F.FONCTIONNAIRE RECR BAC+4 (NC ENST)
6463332	2/3 FONCT B4	2/3 F.FONCTIONNAIRE RECR BAC+4 (NC ENST)
6463333	3/3 FONCT B4	3/3 F.FONCTIONNAIRE RECR BAC+4 (NC ENST)
6463400	F. FONCT B5	F. FONCTIONNAIRE RECR BAC+5 (NC ENST)
7111111	DEA	DEA
7112140	DOCTORAT	DOCTORAT
7112141	1/4 DOCTORAT	DOCTORAT : 1ERE ANNEE
7112142	2/4 DOCTORAT	DOCTORAT : 2EME ANNEE
7112143	3/4 DOCTORAT	DOCTORAT : 3EME ANNEE
7112144	4/4 DOCTORAT	DOCTORAT : 4EME ANNEE
7113100	HDR	HABILITATION A DIRIGER DES RECHERCHES
6419E11	SOCIAL NIV1	DIPLOME SOCIAL DE NIVEAU I
3212A32	1ERE D'ADAPT	1ERE D'ADAPTATION BT BTN
3212B32	1ERE BT	1ERE BT BTA
3213B33	TERM BT-BTA	TERMINALE BT BTA
51C1131	1/3 AUT BAC CAN	BAC CANONIQUE (NC THEOLOGIE) : 1E ANNEE
51C1132	2/3 AUT BAC CAN	BAC CANONIQUE (NC THEOLOGIE) : 2E ANNEE
51C1133	3/3 AUT BAC CAN	BAC CANONIQUE (NC THEOLOGIE) : 3E ANNEE
61C1151	1/5 BAC CAN TH	BAC CANONIQUE THEOLOGIE : 1ERE ANNEE
61C1152	2/5 BAC CAN TH	BAC CANONIQUE THEOLOGIE : 2EME ANNEE
61C1153	3/5 BAC CAN TH	BAC CANONIQUE THEOLOGIE : 3EME ANNEE
61C1154	4/5 BAC CAN TH	BAC CANONIQUE THEOLOGIE : 4EME ANNEE
61C1155	5/5 BAC CAN TH	BAC CANONIQUE THEOLOGIE : 5EME ANNEE
61C2121	1/2 LIC CAN TH	LICENCE CANONIQUE THEOLOGIE : 1ERE ANNEE
61C2122	2/2 LIC CAN TH	LICENCE CANONIQUE THEOLOGIE : 2EME ANNEE
61C2221	1/2 AUT LIC CAN	AUTRE LICENCE CANONIQUE : 1ERE ANNEE
61C2222	2/2 AUT LIC CAN	AUTRE LICENCE CANONIQUE : 2EME ANNEE
6419A11	CAP AVOCAT	CERTIF APTITUDE PROFESSION D'AVOCAT
6419B11	DAF NOTAIRE	DIPLOME APTITUDE FONCTION NOTAIRE
6419C11	DAA	DIPLOME D'AGRONOMIE APPROFONDIE
6419D30	DEO1MM	DIPL ELEVE OFFICIER 1CL MARINE MARCHANDE
71C1141	1/4 DOCT CAN TH	DOCTORAT CANONIQUE THEOLOGIE : 1E ANNNEE
71C1142	2/4 DOCT CAN TH	DOCTORAT CANONIQUE THEOLOGIE : 2E ANNNEE
71C1143	3/4 DOCT CAN TH	DOCTORAT CANONIQUE THEOLOGIE : 3E ANNNEE
71C1144	4/4 DOCT CAN TH	DOCTORAT CANONIQUE THEOLOGIE : 4E ANNNEE
71C1241	1/4 AUT DOC CAN	AUTRE DOCTORAT CANONIQUE : 1ERE ANNEE
71C1242	2/4 AUT DOC CAN	AUTRE DOCTORAT CANONIQUE : 2EME ANNEE
71C1243	3/4 AUT DOC CAN	AUTRE DOCTORAT CANONIQUE : 3EME ANNEE
71C1244	4/4 AUT DOC CAN	AUTRE DOCTORAT CANONIQUE : 4EME ANNEE
A000000	1ER & 2ND DEGRE	ENSEIGNEMENT PRIMAIRE ET SECONDAIRE
B001100	DIVERS SUP	FORMATION POST-SECONDAIRE OU SUPERIEURE

B002000	C2I NIV 1	CERTIFICAT INFORMATIQUE INTERNET 1ER NIV
B002100	C2I NIV 2	CERTIFICAT INFORMATIQUE INTERNET 2ND NIV
B003000	CLES B1	CERTIFICAT DE COMPETENCE LANGUES NIV B1
B003100	CLES B2	CERTIFICAT DE COMPETENCE LANGUES NIV B2
B003200	CLES C1	CERTIFICAT DE COMPETENCE LANGUES NIV C1
B039100	DIV CERT ECOLE	DIVERS CERTIFICATS D'ECOLES
B039200	DIV ATT ECOLE	DIVERSES ATTESTATIONS D'ECOLES
B101100	ENS	FORM DES ECOLES NORMALES SUPERIEURES
B101141	1/4 ENS	FORM ECOLES NORMALES SUPERIEURES : 1E AN
B101142	2/4 ENS	FORM ECOLES NORMALES SUPERIEURES : 2E AN
B101143	3/4 ENS	FORM ECOLES NORMALES SUPERIEURES : 3E AN
B101144	4/4 ENS	FORM ECOLES NORMALES SUPERIEURES : 4E AN
B102100	UNIV PRIVEE	FORM DES ETAB UNIVERSITAIRES PRIVES
B102161	1/6 UNIV PRIVEE	FORM ETAB UNIVERSITAIRES PRIVES : 1E AN
B102162	2/6 UNIV PRIVEE	FORM ETAB UNIVERSITAIRES PRIVES : 2E AN
B102163	3/6 UNIV PRIVEE	FORM ETAB UNIVERSITAIRES PRIVES : 3E AN
B102164	4/6 UNIV PRIVEE	FORM ETAB UNIVERSITAIRES PRIVES : 4E AN
B102165	5/6 UNIV PRIVEE	FORM ETAB UNIVERSITAIRES PRIVES : 5E AN
B102166	6/6 UNIV PRIVEE	FORM ETAB UNIVERSITAIRES PRIVES : 6E AN
B401100	COMMERCE N VISE	FORM ECOLES COMMERCE (DIPL N VISE) VENTE
B402100	ECOLE ART	FORM ECOLES ARTISTIQUES(NC BTS DMA DSAA)
B409100	DIV TEC-PRO SUP	DIVERSES FORM TEC-PRO POST-SECOND OU SUP
B431100	CERT ECOLE COMM	CERTIFICAT ECOLES DE COMMERCE
B431200	ATT ECOLE COMM	ATTESTATION ECOLES DE COMMERCE
B432100	CERT ECOLE ART	CERTIFICAT D'ECOLE D'ARTS
B439100	DIV CER TEC-PRO	DIVERS CERTIFICATS ECOLES TECHNO-PRO
B439200	DIV ATT TEC-PRO	DIVERSES ATTESTATIONS ECOLES TECHNO-PRO
B461100	FORM FONCT	FORMATION DE FONCTIONNAIRE (NC ENST)

ANNEXE 1B. Agrégation *ad hoc* pour définir les classes successives d'inscription (Programme SAS)

Pour suivre les classes successives d'inscription des répondants dans les fichiers des apprenants, la variable FORTRAMS_7, élaborée par la DEPP, est trop détaillée pour nos besoins. Des regroupements ont donc été faits à l'aide du programme SAS présenté ici. Ces choix, suffisant pour le travail exploratoire mené, devraient être réexaminés en cas de nouvel appariement.

```

If source="SCO " then           classe&an.="99-Autres SCO           ";
If source="SIFA" then          classe&an.="99-Autres SIFA          ";

If substr(fortrams7,1,1)= "0" then classe&an.="00-Pré élémentaire ";
If substr(fortrams7,1,1)= "1" then classe&an.="01-Primaire         ";

If substr(fortrams7,1,1)= "2" then classe&an.="07-Niv. Coll-autres ";
If substr(fortrams7,1,6)= "211114" then classe&an.="02-6e          ";
If substr(fortrams7,1,6)= "211124" then classe&an.="03-5e          ";
If substr(fortrams7,1,6)= "211144" then classe&an.="04-4e générale ";
If substr(fortrams7,1,6)= "211154" then classe&an.="05-3e générale ";
If substr(fortrams7,1,4)= "2112" then classe&an.="06-SES/SEGPA      ";

If substr(fortrams7,1,1)= "3" then classe&an.="13-Niv. Lycée-autres ";
If substr(fortrams7,1,4)= "3110" then classe&an.="08-2de générale ";
If substr(fortrams7,1,4)= "3111" then classe&an.="08-2de générale ";
If substr(fortrams7,1,4)= "3112" then classe&an.="09-1re générale ";
If substr(fortrams7,1,4)= "3212" then classe&an.="10-1re techno    ";
If substr(fortrams7,1,4)= "3113" then classe&an.="11-Terminale générale";
If substr(fortrams7,1,4)= "3213" then classe&an.="12-Terminale techno ";

If fortrams7="3311011" then classe&an.="15-CAP TERM          "; *CAP1/1;
If fortrams7="3311021" then classe&an.="14-CAP              "; *CAP1/2;
If fortrams7="3311022" then classe&an.="15-CAP TERM          "; *CAP2/2;
If fortrams7="3311031" then classe&an.="14-CAP              "; *CAP1/3;
If fortrams7="3311032" then classe&an.="14-CAP              "; *CAP2/3;
If fortrams7="3311033" then classe&an.="15-CAP TERM          "; *CAP3/3;

If fortrams7="3312011" then classe&an.="21-BEP TERM          "; *BEP1/1;
If fortrams7="3312021" then classe&an.="20-BEP              "; *BEP1/2;
If fortrams7="3312022" then classe&an.="21-BEP TERM          "; *BEP2/2;

If fortrams7 in ("3313011","3313111") then classe&an.="24-Bac Pro TERM "; *Bac Pro 1/1;
If fortrams7 in ("3313021","3313121") then classe&an.="23-Bac Pro "; *Bac Pro 1/2;
If fortrams7 in ("3313022","3313122") then classe&an.="24-Bac Pro TERM "; *Bac Pro 2/2;
If fortrams7 in ("3313031","3313131") then classe&an.="23-Bac Pro "; *Bac Pro 1/3;
If fortrams7 in ("3313032","3313132") then classe&an.="23-Bac Pro "; *Bac Pro 2/3;
If fortrams7 in ("3313033","3313133") then classe&an.="24-Bac Pro TERM "; *Bac Pro 3/3;

If substr(fortrams7,1,4)="3314" then classe&an.="29-Mention comple. ";

If fortrams7="3315021" then classe&an.="30-Brevet Pro "; *Brevet Pro 1/2;
If fortrams7="3315022" then classe&an.="31-Brevet Pro TERM "; *Brevet Pro 2/2;
If fortrams7="3315031" then classe&an.="30-Brevet Pro "; *Brevet Pro 1/3;
If fortrams7="3315032" then classe&an.="30-Brevet Pro "; *Brevet Pro 2/3;
If fortrams7="3315033" then classe&an.="31-Brevet Pro TERM "; *Brevet Pro 3/3;

If substr(fortrams7,1,1)="4" then classe&an.="35-Fortrams1=4 ";

If substr(fortrams7,1,1)="5" then classe&an.="46-Autre Fortrams1=5 ";
If fortrams7="5141121" then classe&an.="36-CPGE 1/2 ";
If fortrams7="5141122" then classe&an.="37-CPGE 2/2 ";

If fortrams7 in ("5411011","5411111","5411211") then classe&an.="39-BTS TERM ";
* BTS 1/1;
If fortrams7 in ("5411021","5411121","5411221") then classe&an.="38-BTS ";
* BTS 2/1;
If fortrams7 in ("5411022","5411122","5411222") then classe&an.="39-BTS TERM ";
* BTS 2/2;
If fortrams7 in ("5411031","5411131","5411231") then classe&an.="38-BTS ";
* BTS 1/3;
If fortrams7 in ("5411032","5411132","5411232") then classe&an.="38-BTS ";
* BTS 2/3;
If fortrams7 in ("5411033","5411133","5411233") then classe&an.="39-BTS TERM ";

```

```
* BTS 3/3;

If fortrams7="5412121" then classe&an.="44-DUT 1/2      ";
If fortrams7="5412122" then classe&an.="45-DUT 2/2      ";

If substr(fortrams7,1,1)="6" then      classe&an.="58-Autre Fortrams1=6 ";
If substr(fortrams7,1,6)="641121" then classe&an.="47-Licence Pro      ";

If fortrams7="6417121" then classe&an.="48-Master Pro 1/2    ";
If fortrams7="6417122" then classe&an.="49-Master Pro 2/2    ";

If substr(fortrams7,1,4)="6421" then classe&an.="53-Ingenieur  DIV      ";
If fortrams7="6421131" then classe&an.="50-Ingenieur      ";*Ingenieur 1/3;
If fortrams7="6421132" then classe&an.="50-Ingenieur      ";*Ingenieur 2/3;
If fortrams7="6421133" then classe&an.="52-Ingenieur TERM   ";*Ingenieur 3/3;

If substr(fortrams7,1,4)="6422" then classe&an.="57-Commerce  DIV      ";
If fortrams7="6422131" then classe&an.="54-Commerce      ";*Commerce 1/3;
If fortrams7="6422132" then classe&an.="54-Commerce      ";*Commerce 2/3;
If fortrams7="6422133" then classe&an.="56-Commerce TERM   ";*Commerce 3/3;
/ **
```

ANNEXE IC. Répartition des effectifs des fichiers des apprenants transmis par classes, selon notre regroupement *ad hoc*

Ce tableau présente la répartition des effectifs des fichiers des apprenants par classe, selon notre regroupement *ad hoc* fait avec le programme présenté dans l'annexe 1.B, à partir des modalités de la variable FORTRAMS_7 (cf. annexe I.A). La répartition est présentée pour le fichier le plus récent (2016) et pour le fichier le plus ancien (2005).

	Niveau de formation (Niv_fms16)	Effectifs dans le fichier 2016	Effectifs dans le fichier 2005
02-6e	2	2	10 373
03-5e	2	9	8 818
04-4e générale	2	3 622	5 656
05-3e générale	2	11 779	3 306
06-SES/SEGPA	2	1 481	180
07-Niv. Coll-autres	2	22	42
08-2de générale	3	2 567	1 672
09-1re générale	3	1 229	840
10-1re techno	3	1 019	118
11-Terminale générale	3	5 480	659
12-Terminale techno	3	6 238	77
13-Niv. Lycée-autres	3	1 323	9
14-CAP	3	15 583	16
15-CAP TERM	3	35 519	4
20-BEP			109
21-BEP TERM			54
23-Bac pro	3	13 402	12
24-Bac pro TERM	3	35 156	13
29-Mention complèm.	3	1 305	
30-Brevet pro	3	1 724	
31-Brevet pro TERM	3	5 218	
35-Fortrams1=4	4	341	1
36-CPGE 1/2	5	116	58
37-CPGE 2/2	5	869	43
38-BTS	5	5 380	20
39-BTS TERM	5	14 461	16
44-DUT 1/2	5	64	
45-DUT 2/2	5	325	
46-Autre Fortrams1=5	5	475	
47-Licence pro	6	2 189	
48-Master pro 1/2	6	170	
49-Master pro 2/2	6	1 931	
53-Ingenieur DIV	6	1 497	
54-Commerce	6	4	
56-Commerce TERM	6	16	
57-Commerce DIV	6	1 318	
58-Autre Fortrams1=6	6	1 181	
99-Autres SIFA	B	10	
Ensemble		173 025	32 097

Source : DEPP-Fichiers des apprenants 2005 et 2016 restreints à l'échantillon de l'enquête (répondants et non répondants).

Le tableau ci-dessous présente le contenu des lignes « autres » surlignées en gris dans le tableau ci-dessus. Ce contenu est présenté pour l'année 2016, en utilisant le code associé aux 6 premières positions de FORTRAMS_7.

Classe16=13-Niv. Lycée-autres			
fortrams_7	Libellé	Effectif	%
308151	Mission de lutte contre le décrochage scolaire	87	6,6
321113	Seconde techno. Spécifique TMD hôtel	5	0,4
330912	Préparation pré bac	376	28,4
331322	Bac pro. Brevet des métiers d'arts (BMA)	98	7,4
333121	Certificat de spécialisation agricole Niv. 4	96	7,3
338111	Formations d'initiatives locales de Niv. 5	4	0,3
341110	Diplôme paramédical de Niv. 5	165	12,5
341111	Diplôme paramédical de Niv. 5 en un an	123	9,3
341120	Diplôme social de Niv. 5	9	0,7
341122	Diplôme social de Niv. 5	18	1,4
341212	Diplôme social de Niv. 4	11	0,8
343110	Titre homologué de Niv. 5	320	24,2
343120	Titre homologué de Niv. 4	11	0,8
classe16=35-Fortrams1=4			
fortrams_7	Libellé	Effectif	%
414111	Prépa CPGE	3	0,9
440910	Diverses formations techn. et pro. Post 2daire	19	5,6
444111	Diplôme social post secondaire (Niv. 4)	46	13,5
444121	Diplôme social post secondaire (Niv. 4)	273	80,1
classe16=46-Autre Fortrams1=5			
fortrams_7	Libellé	Effectif	%
541133	Diplôme de technicien supérieur (DTS)	99	20,8
541141	Diplôme de conseiller éco. sociale et familiale	111	23,4
541152	Diplôme de métiers d'arts	63	13,3
541313	Diplôme paramédical de Niv. 3	6	1,3
541323	Diplôme social de Niv. 3	126	26,5
541331	Diplôme paramédical Spe. de Niv. 3	4	0,8
541512	DEUST	17	3,6
541611	Diplôme national de technologie spécialisé, DNTS	11	2,3
542912	Divers diplômes d'école de Niv. 3	6	1,3
543510	Titre homologué de Niv. 3	19	4,0
548111	Formations d'initiatives locales de Niv. 3	13	2,7
classe16=46-Autre Fortrams1=6			
fortrams_7	Libellé	Effectif	%
611413	Licence (LMD)	167	14,1
611512	Master (LMD)	165	14,0
612115	Diplôme d'IEP	26	2,2
612413	Diplôme d'établissement, grade de Licence	2	0,2
640510	Divers diplôme techn. et pro. de niveau bac+3	4	0,3
640530	Divers diplôme techn. et pro. de niveau bac+5	5	0,4
641622	Diplôme supérieur d'arts appliqués	37	3,1
641644	Diplôme de paysagiste	2	0,2
641813	Diplôme paramédical de Niv. 2	51	4,3
641913	Diplôme de comptabilité et gestion	423	35,8
641922	Diplôme supérieur de comptabilité et gestion	167	14,1
641962	Diplôme national d'œnologie	2	0,2
64190	?	1	0,1
642923	Divers diplômes d'écoles recrutant à bac+2	1	0,1
642935	Divers diplômes d'écoles recrutant au bac	13	1,1
643525	Certificat d'écoles d'arts recrutant au bac	6	0,5
643710	Titre homologué de Niv. 2	13	1,1

643720	Titre homologué de Niv. 1	96	8,1
classe16=99-Autres SIFA			
fortrams_7	Libellé	Effectif	%
B40110	Formations d'écoles de commerce - Vente	6	60,0
B40210	Formations d'écoles artistiques	4	40,0

Source : DEPP-Fichiers des apprenants 2005 et 2016 restreints à l'échantillon de l'enquête (répondants et non répondants).

ANNEXE 2. Tableaux selon les niveaux de sortie (NIVSOR)

Les tableaux présentés dans cette annexe complètent ceux présents dans le corps du texte en proposant des données selon les niveaux de sortie. Ceux-ci sont définis par le Céreq pour ses analyses. Les niveaux retenus ici correspondent aux deux premières positions de la variable NIVSOR_CEREQ.

Quand le tableau décline un tableau un graphique ou tableau déjà présent dans le corps du texte, sa numérotation reprend celle du corps du texte. Dans le cas contraire, il est identifié par une lettre.

Tableau A • structure des répondants par niveau de sortie, selon les composantes de la base de sondage et proportion d'INE supposé valide

En vert : composantes disposant d'INE « valides ». En rouge, composantes ne disposant pas d'INE « valides ». En blanc, composantes intermédiaires

	Sysca (second degré)	Sifa (appren- tissage)	Universit és	Autres établis- sements	Agricultur e	DREES	Sport	Culture	DGAFP	Cifre	Contrats de prof.		Taux d'INE supposés valides
	SY	SI	UN	G1	AG	G4	SP	CU	G3	CI	CP	Total	
Collège	173	2	2	0	0	0	0	0	0	0	0	177	100%
CAP-BEP-MC de niveau 3	1 546	1 089	0	53	5	145	0	0	0	0	59	2 897	92%
Bac pro	3 382	481	0	42	494	1	0	0	2	0	15	4 417	99%
Autre diplôme de niveau 4	47	356	11	21	0	19	608	0	70	0	52	1 184	35%
Bac techno	530	0	0	0	5	0	0	0	0	0	0	535	100%
Bac général	162	0	0	1	0	0	0	0	0	0	0	163	99%
BTS/DUT	1 567	497	549	185	306	16	0	1	1	0	261	3 383	89%
Autre bac+2	22	5	39	28	0	16	169	0	2	0	24	305	28%
Bac+2 à 4 santé social	33	19	34	36	0	522	0	0	0	0	1	645	46%
Licence pro	0	95	921	75	0	0	0	0	0	0	311	1 402	76%
L3	11	21	1 675	231	0	0	0	8	0	0	22	1 968	97%
Autre bac+3 ou bac+4	9	15	32	248	0	0	14	2	5	0	165	490	25%
M2	1	159	2 724	543	0	0	0	66	42	0	348	3 883	88%
Autre bac+5	7	23	307	200	0	3	0	65	12	1	280	898	56%
Ecole de commerce	0	130	5	618	0	0	0	2	12	0	182	949	63%
Ecole d'ingénieurs	0	190	322	673	0	0	0	2	110	0	38	1 335	89%
Doctorat santé	0	0	189	33	0	0	0	0	10	0	0	232	100%
Doctorat hors santé	0	0	171	63	0	0	0	1	0	66	0	301	76%
	7 490	3 082	6 981	3 050	810	722	791	147	266	67	1 758	25 164	

Champ : répondants (hors post-initiaux).

Note : la pondération associée aux répondants de l'enquête Génération n'est pas utilisée ici.

Source : Céreq-Depp. Rapprochement de l'enquête Génération 2017 à trois ans et des fichiers des apprenants FAERE.

Graphique 3 • Proportion de répondants à Génération 2017 retrouvé dans le fichier FAERE de l'année selon le niveau de sortie

	2016	2015	2014	2013	2012	2011	2010	2009	2008	2007	2006	2005
Collège	97%	95%	92%	86%	28%	6%	3%	3%	2%	1%	1%	1%
CAP-BEP-MC de niveau 3	91%	80%	74%	74%	75%	71%	42%	25%	14%	8%	4%	2%
Bac pro	87%	85%	82%	81%	80%	77%	72%	33%	10%	3%	2%	1%
Autre diplôme de niveau 4	34%	30%	24%	24%	26%	27%	28%	29%	22%	14%	9%	5%
Bac techno	98%	96%	94%	94%	93%	87%	79%	35%	7%	1%	0%	0%
Bac général	98%	95%	95%	94%	94%	93%	93%	46%	10%	1%	1%	0%
BTS/DUT	62%	60%	58%	65%	68%	69%	69%	70%	64%	34%	15%	7%
Autre bac+2	10%	9%	10%	10%	14%	17%	19%	20%	20%	20%	17%	10%
Bac+2 à 4 santé social	8%	7%	7%	12%	22%	29%	31%	33%	34%	34%	28%	18%
Licence pro	32%	26%	30%	42%	52%	57%	60%	62%	63%	63%	36%	18%
L3	3%	24%	39%	57%	71%	78%	81%	83%	74%	62%	40%	24%
Autre bac+3 ou bac+4	6%	6%	6%	10%	12%	15%	17%	18%	19%	19%	14%	10%
M2	15%	6%	4%	7%	12%	34%	53%	64%	68%	69%	70%	68%
Autre bac+5	4%	3%	2%	4%	7%	18%	28%	35%	40%	43%	45%	46%
Ecole de commerce	14%	5%	1%	5%	18%	31%	39%	42%	44%	44%	43%	43%
Ecole d'ingénieurs	17%	16%	16%	29%	37%	59%	68%	72%	72%	73%	72%	71%
Doctorat santé	0%	0%	0%	0%	0%	4%	9%	14%	19%	22%	37%	58%
Doctorat hors santé	0%	0%	0%	0%	0%	1%	3%	8%	16%	26%	34%	41%

Champ : répondants (hors post-initiaux).

Note : la pondération associée aux répondants de l'enquête Génération n'est pas utilisée ici.

Source : Céreq-Depp. Rapprochement de l'enquête Génération 2017 à trois ans et des fichiers des apprenants FAERE.

Graphiques 4 et 5, Tableau 7 • Nombre d'années de présence dans les fichiers annuels 2005 à 2016, selon le niveau de sortie

	% présent au moins une fois	Nombre de fichiers annuels dans lesquels les répondants retrouvés au moins une fois sont présents												
		1	2	3	4	5	6	7	8	9	10	11	12	Total
Collège	99%	2%	3%	7%	59%	23%	2%	1%	1%	1%	0%	0%	0%	100%
CAP-BEP-MC de niveau 3	91%	5%	3%	3%	3%	9%	35%	20%	12%	6%	2%	1%	0%	100%
Bac professionnel	88%	1%	2%	1%	2%	5%	8%	45%	25%	8%	1%	0%	0%	100%
Autre diplôme de niveau 4	34%	5%	6%	2%	1%	2%	2%	4%	24%	31%	13%	8%	2%	100%
Bac technologique	98%	2%	2%	0%	2%	6%	8%	45%	29%	6%	0%	0%	0%	100%
Bac général	98%	2%	1%	1%	1%	1%	1%	49%	37%	9%	0%	0%	0%	100%
BTS/DUT	78%	2%	3%	1%	1%	2%	4%	13%	12%	37%	18%	5%	2%	100%
Autre bac+2	24%	3%	3%	5%	7%	3%	7%	19%	15%	12%	18%	8%	0%	100%
Bac+2 à 4 santé social	38%	2%	3%	5%	4%	4%	8%	44%	12%	7%	5%	4%	2%	100%
Licence professionnelle	69%	4%	1%	1%	3%	4%	4%	20%	17%	20%	18%	7%	1%	100%
Licence générale	89%	1%	1%	2%	3%	4%	5%	50%	21%	9%	3%	1%	0%	100%
Autre bac+3 ou bac+4	21%	1%	5%	4%	6%	9%	6%	27%	13%	13%	11%	7%	1%	100%
Master	75%	3%	2%	4%	6%	13%	22%	30%	9%	7%	2%	1%	1%	100%
Autre bac+5	49%	5%	5%	9%	12%	15%	20%	21%	6%	4%	1%	1%	3%	100%
Ecole de commerce	50%	4%	5%	2%	4%	7%	14%	24%	22%	14%	3%	1%	0%	100%
Ecole d'ingénieurs	79%	2%	2%	3%	2%	3%	11%	24%	13%	27%	7%	2%	4%	100%
Doctorat santé	59%	36%	26%	7%	7%	10%	7%	7%	0%	0%	0%	0%	0%	100%
Doctorat hors santé	42%	18%	18%	26%	18%	12%	6%	2%	0%	0%	0%	0%	0%	100%

Champ : répondants (hors post-initiaux).

Note : la pondération associée aux répondants de l'enquête Génération n'est pas utilisée ici.

Source : Céreq-Depp. Rapprochement de l'enquête Génération 2017 à trois ans et des fichiers des apprenants FAERE.

Tableau 8 • Première observée pour les répondants à Génération 2017 retrouvés au moins une fois dans les fichiers des apprenants

1ère année observée pour les répondants à Génération 2017 retrouvés au moins une fois dans les fichiers des apprenants												
Selon le niveau de sortie	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
Collège	1%	0%	0%	1%	1%	1%	3%	22%	59%	6%	3%	2%
CAP-BEP-MC de niveau 3	3%	2%	4%	7%	12%	19%	33%	7%	2%	3%	3%	5%
Bac pro	1%	1%	2%	8%	26%	45%	7%	5%	1%	1%	2%	1%
Autre diplôme de niveau 4	14%	10%	15%	26%	20%	0%	0%	1%	1%	1%	5%	5%
Bac techno	0%	0%	0%	6%	29%	45%	8%	6%	2%	0%	2%	2%
Bac général	0%	1%	0%	10%	36%	48%	1%	1%	1%	1%	1%	2%
BTS/DUT	9%	10%	25%	39%	8%	1%	2%	1%	1%	1%	3%	2%
Autre bac+2	44%	26%	15%	3%	5%	0%	0%	1%	0%	1%	1%	3%
Bac+2 à 4 santé social	47%	27%	17%	2%	0%	2%	1%	0%	0%	2%	1%	2%
Licence pro	26%	27%	38%	1%	0%	1%	1%	1%	0%	1%	0%	4%
L3	27%	19%	25%	14%	11%	1%	1%	1%	0%	0%	0%	0%
Autre bac+3 ou bac+4	48%	20%	24%	1%	0%	1%	1%	0%	0%	1%	4%	0%
M2	91%	3%	1%	1%	1%	0%	0%	0%	0%	0%	0%	2%
Autre bac+5	93%	2%	1%	1%	1%	1%	0%	0%	0%	0%	0%	0%
Ecole de commerce	87%	1%	2%	1%	1%	1%	0%	1%	0%	0%	2%	3%
Ecole d'ingénieurs	91%	2%	1%	1%	1%	1%	1%	0%	0%	1%	0%	1%
Doctorat santé	99%	0%	1%	1%	0%	0%	0%	0%	0%	0%	0%	0%
Doctorat hors santé	99%	0%	0%	0%	0%	0%	0%	1%	0%	0%	0%	0%

Champ : répondants (hors post-initiaux).

Note : la pondération associée aux répondants de l'enquête Génération n'est pas utilisée ici.

Source : Céreq-Depp. Rapprochement de l'enquête Génération 2017 à trois ans et des fichiers des apprenants FAERE.

Tableau 9 • Dernière année observée pour les répondants à Génération 2017 retrouvés au moins une fois dans les fichiers des apprenants

Dernière année observée pour les répondants à Génération 2017 retrouvés au moins une fois dans les fichiers des apprenants												
Selon le niveau de sortie	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
Collège	0%	0%	0%	1%	0%	1%	0%	0%	0%	0%	1%	98%
CAP-BEP-MC de niveau 3	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	100%
Bac pro	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	1%	99%
Autre diplôme de niveau 4	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	99%
Bac techno	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	100%
Bac général	0%	0%	0%	0%	0%	0%	0%	1%	0%	0%	0%	99%
BTS/DUT	0%	0%	0%	0%	1%	1%	1%	2%	4%	8%	3%	80%
Autre bac+2	0%	1%	1%	5%	4%	5%	7%	15%	10%	7%	1%	42%
Bac+2 à 4 santé social	0%	1%	0%	2%	4%	7%	16%	25%	22%	1%	0%	22%
Licence pro	0%	0%	0%	0%	2%	4%	4%	7%	14%	4%	18%	46%
L3	0%	0%	1%	2%	3%	4%	8%	16%	21%	17%	24%	3%
Autre bac+3 ou bac+4	1%	0%	2%	4%	9%	7%	9%	11%	21%	5%	6%	27%
M2	1%	1%	2%	5%	12%	21%	25%	7%	5%	1%	1%	19%
Autre bac+5	4%	4%	7%	11%	15%	20%	21%	5%	3%	1%	0%	9%
Ecole de commerce	0%	1%	1%	3%	4%	13%	22%	20%	7%	1%	0%	28%
Ecole d'ingénieurs	0%	0%	0%	1%	2%	10%	23%	11%	28%	2%	0%	22%
Doctorat santé	36%	26%	7%	7%	9%	7%	7%	0%	0%	0%	1%	1%
Doctorat hors santé	18%	18%	25%	18%	11%	6%	2%	1%	0%	1%	1%	0%

Champ : répondants (hors post-initiaux).

Note : la pondération associée aux répondants de l'enquête Génération n'est pas utilisée ici.

Source : Céreq-Depp. Rapprochement de l'enquête Génération 2017 à trois ans et des fichiers des apprenants FAERE.

Tableau 10 • Nombre d'années « vides » entre la première et la dernière année observée pour les répondants à Génération 2017 retrouvés au moins une fois dans les fichiers des apprenants

	Nombre d'années "manquantes" entre la première année et la dernière année d'observation									
Selon le niveau de sortie	0	1	2	3	4	5	6	7	8	9
Collège	99%	1%	0%	0%	0%	0%	0%	0%	0%	0%
CAP-BEP-MC de niveau 3	85%	7%	4%	2%	1%	1%	0%	0%	0%	0%
Bac pro	94%	3%	2%	1%	0%	0%	0%	0%	0%	0%
Autre diplôme de niveau 4	66%	15%	9%	4%	3%	2%	0%	1%	0%	0%
Bac techno	98%	2%	1%	0%	0%	0%	0%	0%	0%	0%
Bac général	98%	1%	0%	1%	0%	0%	0%	0%	0%	0%
BTS/DUT	84%	9%	3%	2%	1%	0%	0%	0%	0%	0%
Autre Bac+2	81%	8%	5%	4%	1%	0%	0%	0%	0%	0%
Bac+2 à 4 santé social	88%	4%	2%	1%	2%	1%	1%	1%	0%	0%
Licence pro	68%	7%	12%	7%	2%	2%	1%	0%	0%	0%
L3	96%	2%	1%	0%	0%	0%	0%	0%	0%	0%
Autre bac+3 ou bac+4	86%	5%	5%	3%	1%	1%	0%	0%	0%	0%
M2	81%	2%	2%	3%	5%	4%	2%	1%	0%	0%
Autre bac+5	94%	2%	1%	1%	1%	1%	0%	0%	0%	0%
Ecole de commerce	74%	3%	3%	10%	3%	5%	2%	0%	0%	0%
Ecole d'ingénieurs	82%	4%	7%	3%	2%	1%	0%	0%	0%	0%
Doctorat santé	98%	1%	0%	0%	0%	1%	1%	0%	0%	0%
Doctorat hors santé	98%	1%	0%	0%	0%	0%	1%	0%	1%	0%

Champ : répondants (hors post-initiaux).

Note : la pondération associée aux répondants de l'enquête Génération n'est pas utilisée ici.

Source : Céreq-Depp. Rapprochement de l'enquête Génération 2017 à trois ans et des fichiers des apprenants FAERE.

Tableau 12 • Repérage des inscriptions en classe de collège et de lycée

	Proportion de répondants à Génération 2017 pour lesquels la classe considérée est retrouvée parmi les répondants retrouvés au moins une fois dans les fichiers de la DEPP							% de repérés au moins une fois dans les fichiers FAERE
	6e	5e	4e	3e	2de	1ère	Term.	
Collège	71%	70%	72%	74%	2%	2%	2%	99%
CAP-BEP-MC de niveau 3	63%	63%	62%	61%	6%	4%	3%	91%
Bac professionnel	88%	88%	88%	88%	8%	2%	1%	88%
Autre diplôme de niveau 4	75%	79%	79%	79%	26%	24%	23%	34%
Bac technologique	92%	93%	94%	95%	95%	90%	74%	98%
Bac général	95%	94%	96%	96%	98%	98%	100%	98%
BTS/DUT	84%	86%	88%	90%	58%	57%	56%	78%
Autre bac+2	63%	71%	77%	86%	71%	77%	77%	24%
Bac+2 à 4 santé social	67%	77%	83%	86%	84%	91%	92%	38%
Licence professionnelle	78%	84%	88%	91%	82%	82%	82%	69%
Licence générale	77%	85%	90%	94%	85%	87%	87%	89%
Autre bac+3 ou bac+4	58%	73%	85%	88%	86%	88%	88%	21%
Master	31%	58%	78%	87%	91%	94%	95%	75%
Autre bac+5	20%	41%	61%	76%	85%	90%	96%	49%
Ecole de commerce	18%	59%	77%	86%	89%	92%	92%	50%
Ecole d'ingénieurs	46%	75%	88%	91%	91%	93%	93%	79%
Doctorat santé	0%	4%	15%	26%	37%	63%	98%	59%
Doctorat hors santé	1%	1%	1%	18%	46%	70%	86%	42%

Champ : répondants (hors post-initiaux).

Note : la pondération associée aux répondants de l'enquête Génération n'est pas utilisée ici.

Source : Céreq-Depp. Rapprochement de l'enquête Génération 2017 à trois ans et des fichiers des apprenants FAERE.

Tableau 13 • Poursuite du parcours scolaire

	Parmi les répondants à Génération 2017 retrouvés au moins une fois dans les fichiers de la DEPP, proportion de répondants ayant au moins une année d'inscription en ...					% de repérés au moins une fois dans les fichiers FAERE
	CAP	BEP	Bac pro	Classe préparatoire	BTS	
Collège	3%	0%	1%	0%	0%	99%
CAP-BEP-MC de niveau 3	95%	0%	17%	0%	2%	91%
Bac professionnel	15%	0%	98%	0%	0%	88%
Autre diplôme de niveau 4	51%	3%	23%	0%	4%	34%
Bac technologique	1%	0%	2%	0%	0%	98%
Bac général	0%	0%	0%	0%	0%	98%
BTS/DUT	3%	1%	37%	1%	78%	78%
Autre bac+2	7%	4%	12%	7%	10%	24%
Bac+2 à 4 santé social	0%	5%	2%	0%	15%	38%
Licence professionnelle	1%	1%	9%	2%	44%	69%
Licence générale	1%	1%	12%	4%	12%	89%
Autre bac+3 ou bac+4	0%	0%	6%	6%	21%	21%
Master	0%	1%	1%	11%	9%	75%
Autre bac+5	0%	1%	1%	16%	8%	49%
Ecole de commerce	0%	1%	0%	37%	8%	50%
Ecole d'ingénieurs	0%	1%	1%	48%	7%	79%
Doctorat santé	0%	0%	0%	14%	1%	59%
Doctorat hors santé	0%	0%	1%	40%	2%	42%

Champ : répondants (hors post-initiaux).

Note : la pondération associée aux répondants de l'enquête Génération n'est pas utilisée ici.

Source : Céreq-Depp. Rapprochement de l'enquête Génération 2017 à trois ans et des fichiers des apprenants FAERE.

Tableau 14 • Première classe retrouvée dans les données administratives mobilisées

	Classe d'inscription lors de la première année de présence dans les fichiers FAERE 2005 à 2016 (pour les répondants à Génération 2017 retrouvés au moins une fois dans ces fichiers)											% de repérés au moins une fois dans les fichiers FAERE
	6e	5e	4e	3e	SES/SEGPA	Autres niv. Collège	2de	1ère	Terminale	Autres niv. Lycée	Autres (>)	
Collège	71%	2%	2%	4%	17%	5%	0%	0%	0%	0%	0%	99%
CAP-BEP-MC de niveau 3	63%	2%	2%	2%	17%	4%	0%	0%	0%	2%	8%	91%
Bac pro	88%	1%	1%	1%	3%	0%	0%	0%	0%	1%	4%	88%
Autre diplôme de niveau 4	75%	5%	2%	2%	1%	0%	1%	0%	0%	1%	12%	34%
Bac techno	92%	1%	1%	1%	0%	0%	2%	1%	0%	0%	2%	98%
Bac général	95%	0%	1%	0%	0%	0%	1%	1%	2%	0%	0%	98%
BTS/DUT	84%	3%	2%	2%	0%	0%	1%	0%	0%	0%	7%	78%
Autre bac+2	63%	10%	4%	10%	0%	0%	3%	3%	0%	0%	8%	24%
Bac+2 à 4 santé social	67%	10%	7%	3%	0%	0%	5%	1%	0%	0%	7%	38%
Licence pro	78%	7%	4%	3%	0%	0%	1%	1%	0%	0%	6%	69%
L3	77%	8%	5%	4%	0%	0%	2%	1%	1%	0%	1%	89%
Autre bac+3 ou bac+4	58%	15%	12%	3%	0%	0%	5%	2%	1%	0%	5%	21%
M2	31%	27%	20%	9%	0%	0%	6%	2%	1%	0%	4%	75%
Autre bac+5	20%	21%	20%	15%	0%	0%	10%	6%	5%	0%	2%	49%
Ecole de commerce	18%	41%	18%	8%	0%	0%	5%	1%	0%	0%	8%	50%
Ecole d'ingénieurs	46%	30%	13%	3%	0%	0%	2%	1%	0%	0%	5%	79%
Doctorat santé	0%	4%	11%	11%	0%	0%	11%	26%	36%	0%	1%	59%
Doctorat hors santé	1%	0%	0%	17%	0%	0%	30%	24%	17%	0%	12%	42%

Champ : répondants (hors post-initiaux).

Note : la pondération associée aux répondants de l'enquête Génération n'est pas utilisée ici.

Source : Céreq-Depp. Rapprochement de l'enquête Génération 2017 à trois ans et des fichiers des apprenants FAERE.

Tableau 15 • Dernière classe retrouvée dans les données administratives mobilisées

	Classe d'inscription lors de la dernière année de présence dans les fichiers FAERE 2005 à 2016 (pour les répondants à Génération 2017 retrouvés au moins une fois dans ces fichiers)											% de repérés au moins une fois dans les fichiers FAERE
	Classe de 6e à 1ère (ie SES/SEGPA)	Terminale	CAP (année terminale ou non)	Bac pro (année terminale ou non)	Brevet pro (année terminale ou non)	CGPE	BTS (année terminale ou non)	DUT (année terminale ou non)	Licence pro	Master pro (année terminale ou non)	Autres	
Collège	95%	1%	3%	0%	0%	0%	0%	0%	0%	0%	0%	99%
CAP-BEP-MC de niveau 3	0%	0%	93%	0%	0%	0%	0%	0%	0%	0%	6%	91%
Bac pro	1%	0%	0%	98%	1%	0%	0%	0%	0%	0%	0%	88%
Autre diplôme de niveau 4	0%	2%	0%	0%	81%	0%	0%	0%	0%	0%	17%	34%
Bac techno	26%	74%	0%	0%	0%	0%	0%	0%	0%	0%	0%	98%
Bac général	0%	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	98%
BTS/DUT	0%	17%	0%	1%	0%	1%	78%	2%	0%	0%	1%	78%
Autre bac+2	0%	34%	3%	4%	0%	3%	5%	0%	0%	0%	51%	24%
Bac+2 à 4 santé social	1%	64%	0%	1%	0%	0%	4%	0%	0%	0%	29%	38%
Licence pro	2%	27%	0%	1%	0%	1%	22%	0%	46%	0%	1%	69%
L3	1%	73%	0%	9%	0%	1%	11%	0%	0%	0%	5%	89%
Autre bac+3 ou bac+4	2%	55%	0%	1%	0%	1%	9%	0%	0%	0%	33%	21%
M2	1%	62%	0%	0%	0%	3%	6%	0%	1%	17%	11%	75%
Autre bac+5	1%	68%	0%	0%	0%	5%	4%	0%	0%	1%	20%	49%
Ecole de commerce	0%	36%	0%	0%	0%	1%	4%	0%	0%	0%	57%	50%
Ecole d'ingénieurs	1%	31%	0%	0%	0%	3%	1%	0%	0%	0%	63%	79%
Doctorat santé	1%	83%	0%	0%	0%	2%	1%	0%	0%	1%	12%	59%
Doctorat hors santé	1%	54%	0%	1%	0%	6%	1%	0%	0%	2%	35%	42%

Champ : répondants (hors post-initiaux).

Note : la pondération associée aux répondants de l'enquête Génération n'est pas utilisée ici.

Source : Céreq-Depp. Rapprochement de l'enquête Génération 2017 à trois ans et des fichiers des apprenants FAERE.

Tableau 16 • Redoublements repérés dans les fichiers FAERE

Non pondéré	Proportion de répondants à Génération 2017 concernée parmi ceux retrouvés au moins une fois dans les fichiers de la DEPP			% de repérés au moins une fois dans les fichiers FAERE
	Redoublement au collège	Redoublement au Lycée	A été en classe de collège SEGPA ou assimilé	
Collège	26%	1%	23%	99%
CAP-BEP-MC de niveau 3	14%	1%	22%	91%
Bac professionnel	18%	1%	4%	88%
Autre diplôme de niveau 4	13%	7%	2%	34%
Bac technologique	13%	37%	0%	98%
Bac général	8%	43%	0%	98%
BTS/DUT	10%	14%	0%	78%
Autre bac+2	3%	21%	0%	24%
Bac+2 à 4 santé social	2%	17%	0%	38%
Licence professionnelle	5%	16%	0%	69%
Licence générale	8%	22%	0%	89%
Autre bac+3 ou bac+4	7%	17%	0%	21%
Master	2%	12%	0%	75%
Autre bac+5	1%	10%	0%	49%
Ecole de commerce	2%	12%	0%	50%
Ecole d'ingénieurs	1%	6%	0%	79%
Doctorat santé	0%	2%	0%	59%
Doctorat hors santé	0%	2%	0%	42%

Champ : répondants (hors post-initiaux).

Note : la pondération associée aux répondants de l'enquête Génération n'est pas utilisée ici.

Source : Céreq-Depp. Rapprochement de l'enquête Génération 2017 à trois ans et des fichiers des apprenants FAERE.

Tableau 17 • Etre redoublant au collège dans l'enquête et les données administratives

	Sur tous les répondants à Génération 2017			Sur tous les répondants à Génération retrouvés au moins une fois dans les fichiers FAERE		% de repérés au moins une fois dans les fichiers FAERE
	% des redoublants dans l'enquête repérés comme tels dans FAERE	% des non redoublants dans l'enquête repérés comme redoublants dans FAERE	% des hors champs de la question à l'enquête repérés comme redoublants dans FAERE	% des redoublants dans l'enquête repérés comme tels dans FAEREE	% des non redoublants dans l'enquête repérés comme redoublants dans FAERE	
Collège						99%
CAP-BEP-MC de niveau 3	58,7%	1,5%	0,0%	68%	1,7%	91%
Bac pro	71,1%	1,4%	100,0%	84%	1,6%	88%
Autre diplôme de niveau 4	37,1%	0,1%	0,0%	76%	0,2%	34%
Bac techno	77,6%	1,3%	s.o.	79%	1,3%	98%
Bac général	86,1%	0,0%	s.o.	86%	0,0%	98%
BTS/DUT	58,8%	0,5%	0,0%	80%	0,6%	78%
Autre bac+2	31,7%	0,0%	0,0%	67%	0,0%	24%
Bac+2 à 4 santé social	8,2%	0,0%	0,0%	33%	0,0%	38%
Licence pro	55,7%	0,1%	0,0%	78%	0,1%	69%
L3	67,1%	0,5%	11,1%	78%	0,6%	89%
Autre bac+3 ou bac+4						21%
M2						75%
Autre bac+5						49%
Ecole de commerce						50%
Ecole d'ingénieurs						79%
Doctorat santé						59%
Doctorat hors santé						42%

Champ : répondants (hors post-initiaux).

Note : la pondération associée aux répondants de l'enquête Génération n'est pas utilisée ici.

La question du redoublement en collège (PS050) n'est pas posée aux jeunes inscrits en 2016/2017 dans une formation de niveau bac+4 et plus, ou dans une classe antérieure à la troisième.

Source : Céreq-Depp. Rapprochement de l'enquête Génération 2017 à trois ans et des fichiers des apprenants FAERE.

Tableau 19 • Cohérence/Complémentarité de l'enquête et des données administratives sur l'inscription en classe de troisième

			TOUS INDIVIDUS		CHAMP = AU MOINS RETROUVES UNE FOIS DANS FAERE	
	% des répondants déclarant avoir fait une 3e générale dans Génération	% de répondants de Génération à qui PS040 n'a pas été posée	% répondants ayant une 3e générale repérée dans FAERE parmi ceux déclarant n'avoir pas fait de 3e dans Génération	% répondants sans 3e générale repérée dans FAERE parmi ceux déclarant avoir fait une 3e générale dans Génération	% répondants ayant une 3e générale repérée dans FAERE parmi ceux déclarant n'avoir pas fait de 3e dans Génération	% répondants sans 3e générale repérée dans FAERE parmi ceux déclarant avoir fait une 3e générale dans Génération
Collège		94,3%				
CAP-BEP-MC de niveau 3	58,7%	0,3%	15,9%	32,9%	16,5%	17,7%
Bac pro	73,7%	0,4%	8,4%	17,1%	9,7%	6,2%
Autre diplôme de niveau 4	84,5%	3,6%	22,0%	59,5%	68,9%	19,1%
Bac techno	95,8%	0,0%	22,1%	6,7%	22,1%	5,5%
Bac général	99,3%	0,0%	s.o.	6,7%	s.o.	5,4%
BTS/DUT	93,5%	0,7%	66,3%	28,9%	66,3%	9,2%
Autre bac+2	85,9%	9,0%	s.o.	61,4%	s.o.	17,2%
Bac+2 à 4 santé social	97,1%	1,2%	s.o.	67,8%	s.o.	16,6%
Licence pro	95,0%	2,0%	s.o.	29,8%	s.o.	9,1%
L3	92,4%	3,4%	s.o.	11,5%	s.o.	5,4%
Autre bac+3 ou bac+4		87,4%				
M2		99,2%				
Autre bac+5		97,2%				
Ecole de commerce		100,0%				
Ecole d'ingénieurs		100,0%				
Doctorat santé		100,0%				
Doctorat hors santé		100,0%				

Champ : répondants (hors post-initiaux).

Note : la pondération associée aux répondants de l'enquête Génération n'est pas utilisée ici.

La question de l'inscription en troisième (PS040) n'est pas posée aux jeunes inscrits en 2016/2017 dans une formation de niveau bac+4 et plus, ou dans une classe antérieure à la troisième.

Source : Céreq-Depp. Rapprochement de l'enquête Génération 2017 à trois ans et des fichiers des apprenants FAERE.

Tableau 20 • Cohérence/Complémentarité de l'enquête et des données administratives sur l'orientation après une troisième générale

Classe post troisième identifiée dans FAERE	Orientation après la troisième selon l'enquête Génération (PS060)									
	Seconde générale ou techno	Première année de CAP	Seconde pro	Pas de classe après la troisième	Seconde à l'étranger	1ère année de BEP	Autre	Question non posée	%	Total
Collège								9 986	94%	10 591
CAP-BEP-MC de niveau 3	17 771	52 049	28 482	1 101	972	1 300	620	1 404	1%	103 698
Bac pro	13 613	15 191	86 570	353	466	855	181	683	1%	117 913
Autre diplôme de niveau 4	16 237	7 756	7 438	35	382	47	762	316	1%	32 974
Bac techno	23 839	90	327	16	0	140	0	76	0%	24 489
Bac général	15 587	0	0	85	0	0	0	0	0%	15 672
BTS/DUT	72 564	2 099	29 199	5	668	145	921	842	1%	106 443
Autre bac+2	4 550	56	563	0	148	5	109	535	9%	5 966
Bac+2 à 4 santé social	21 588	88	964	0	111	0	844	287	1%	23 882
Licence pro	26 668	197	3 513	33	302	46	98	642	2%	31 498
L3	66 116	608	5 750	23	1 603	55	318	2 839	4%	77 311
Autre bac+3 ou bac+4								12 025	87%	13 790
M2								94 215	99%	95 482
Autre bac+5								25 703	96%	26 754
Ecole de commerce								22 612	100%	22 612
Ecole d'ingénieurs								22 224	100%	22 224
Doctorat santé								6 862	100%	6 862
Doctorat hors santé								7 841	100%	7 841

Champ : répondants (hors post-initiaux).

Note : la pondération associée aux répondants de l'enquête Génération n'est pas utilisée ici.

La question de l'inscription en 3ème (PS060) n'est pas posée aux jeunes inscrits en 2016/2017 dans une formation de niveau bac+4 et plus, ou dans une classe antérieure à la troisième.

Source : Céreq-Depp. Rapprochement de l'enquête Génération 2017 à trois ans et des fichiers des apprenants FAERE.

Tableau 22 • Inscriptions aux examens repérées dans les fichiers FAERE 2013 à 2016 pour les répondants de Génération 2017

Non pondéré	Proportions de personnes concernées parmi les répondants à l'enquête Génération 2017 repérés au moins une fois dans les fichiers FAERE 2005 à 2016						% de repérés au moins une fois dans les fichiers FAERE
	Brevet		Bac		CAP/BEP		
	1 fois	2 fois ou plus	1 fois	2 fois ou plus	1 fois	2 fois ou plus	
Collège	54%	6%	2%	0%	1%	1%	99%
CAP-BEP-MC de niveau 3	54%	3%	12%	1%	61%	18%	91%
Bac professionnel	80%	7%	68%	8%	75%	9%	88%
Autre diplôme de niveau 4	77%	2%	38%	3%	55%	10%	34%
Bac technologique	90%	4%	63%	10%	1%	0%	98%
Bac général	94%	2%	73%	24%	0%	0%	98%
BTS/DUT	87%	3%	85%	6%	33%	2%	78%
Autre bac+2	86%	0%	84%	5%	16%	0%	24%
Bac+2 à 4 santé social	86%	0%	89%	5%	6%	0%	38%
Licence professionnelle	90%	1%	86%	4%	9%	0%	69%
Licence générale	91%	2%	90%	9%	11%	1%	89%
Autre bac+3 ou bac+4	88%	0%	89%	4%	6%	0%	21%
Master	87%	0%	92%	4%	1%	0%	75%
Autre bac+5	75%	0%	95%	2%	1%	0%	49%
Ecole de commerce	85%	0%	91%	1%	1%	0%	50%
Ecole d'ingénieurs	90%	0%	92%	1%	1%	0%	79%
Doctorat santé	26%	0%	98%	0%	0%	0%	59%
Doctorat hors santé	18%	0%	86%	1%	1%	0%	42%

Champ : répondants (hors post-initiaux).

Note : la pondération associée aux répondants de l'enquête Génération n'est pas utilisée ici.

Source : Céreq-Depp. Rapprochement de l'enquête Génération 2017 à trois ans et des fichiers des apprenants FAERE.

Tableau 26 • Cohérence du type de baccalauréat déclaré à l'enquête avec celui repéré dans les données administratives

	Réussite au moins une fois au bac (tous types) avec les fichiers FAERE		Total des bacheliers dans l'enquête Génération	dont % repérés comme tels avec les fichiers FAERE	Cohérence de l'année et du type du bac pour les répondants à l'enquête repéré avec un bac avec les fichiers FAERE		
	Non	Oui			Non	Oui	Taux de cohérence
Collège	0	3	3	100%	1	2	66,7%
CAP-BEP-MC de niveau 3	297	269	566	48%	67	202	75,1%
Bac professionnel	558	2 344	2 902	81%	30	2296	98,7%
Autre diplôme de niveau 4	626	133	759	18%	25	108	81,2%
Bac technologique	10	309	319	97%	0	309	100,0%
Bac général	7	98	105	93%	0	98	100,0%
BTS/DUT	959	2 406	3 365	72%	486	1914	79,8%
Autre bac+2	222	65	287	23%	11	54	83,1%
Bac+2 à 4 santé social	406	230	636	36%	39	191	83,0%
Licence professionnelle	528	872	1 400	62%	383	489	56,1%
Licence générale	243	1 717	1 960	88%	431	1285	74,9%
Autre bac+3 ou bac+4	389	97	486	20%	28	69	71,1%
Master	1 090	2 767	3 857	72%	186	2581	93,3%
Autre bac+5	462	429	891	48%	32	397	92,5%
Ecole de commerce	517	431	948	45%	20	411	95,4%
Ecole d'ingénieurs	354	977	1 331	73%	45	928	95,4%
Doctorat santé	98	134	232	58%	11	123	91,8%
Doctorat hors santé	191	108	299	36%	6	102	94,4%

Champ : répondants (hors post-initiaux).

Note : la pondération associée aux répondants de l'enquête Génération n'est pas utilisée ici.

Source : Céreq-Depp. Rapprochement de l'enquête Génération 2017 à trois ans et des fichiers des apprenants FAERE.

Tableau 28 • Nombre d'années d'inscription repérées dans les fichiers des apprenants 2005 à 2016, par secteur public/privé

	Nombre moyen d'années d'inscription retrouvés dans les fichiers FAERE		
	Ensemble	Dans le secteur privé	Dans le secteur public
Collège	4,2	0,4	3,8
CAP-BEP-MC de niveau 3	6,2	1,3	4,8
Bac pro	7,0	1,0	6,0
Autre diplôme de niveau 4	7,9	3,4	4,5
Bac techno	6,9	0,6	6,3
Bac général	7,3	0,6	6,7
BTS/DUT	8,2	1,6	6,6
Autre bac+2	7,4	1,6	5,8
Bac+2 à 4 santé social	6,9	2,2	4,7
Licence pro	7,8	1,5	6,3
L3	7,1	1,2	5,9
Autre bac+3 ou bac+4	7,2	1,8	5,3
M2	6,2	1,3	4,9
Autre bac+5	5,6	1,3	4,3
Ecole de commerce	6,7	2,7	4,0
Ecole d'ingénieurs	7,7	2,0	5,7
Doctorat santé	2,8	0,8	2,0
Doctorat hors santé	3,1	0,4	2,7

Champ : répondants (hors post-initiaux).

Note : la pondération associée aux répondants de l'enquête Génération n'est pas utilisée ici.

Source : Céreq-Depp. Rapprochement de l'enquête Génération 2017 à trois ans et des fichiers des apprenants FAERE.

ANNEXE 3. Programme SAS

```
/* **** */
/* **** */
/* **** */
/* Rapprochement expérimental entre les données de la DEPP
/* et Génération 2017 (SJ - début le 30/08/2021)
/* **** */
/* **** */
/* **** */
/* répertoire de sortie */
libname sor "P:\_Géné-Faere-DEPP\Données sortie";run;
/* répertoire des données DEPP */
libname depp "P:\_Géné-Faere-DEPP\Données DEPP";run;

/* répertoire de Génération2004 à 7 ans - table g047ansindividus8sspi */
libname gene "U:\DEEVA\03-ENQUETES GENERATION\02-bases de données\GENE2017\04 - Bases et documentations définitives";
run;

Title "Contenu des tables FAERE";
Proc contents data=depp.generationfaere16;run;
Proc contents data=depp.generationfaere15;run;
Proc contents data=depp.generationfaere11;run;
Proc contents data=depp.generationfaere13;run;
Proc contents data=depp.generationfaere12;run;
Proc contents data=depp.generationfaere06;run;

Title "Contenu des tables de l'enquête Génération 2017";
Proc contents data=gene.g17individus;run;
Proc contents data=gene.g17calact;run;

/* **** */
```

```

/*****
/*****
/***** 0.EXAMEN DE LA TABLE DE PASSAGE INE/IDENT TRANSMISE A LA DEPP */
/*****
/*****
/*****
/*****

* 0.a) Verif unicité des IDENT dans la table de passage INE-IDENT transmise à DEPP;
Proc sort data=depp.ine_gene2017; by ident;run;
Data verif; set depp.ine_gene2017;
Retain n 0 ident1;
If ident=ident1 then n=n+1;
Else n=1;
Ident1=ident;
PrefixIDENT=substr(IDENT,1,2);
run;
Data verif2; set verif; by ident; if last.ident;run;
Title "Il y a t'il des IDENT en doublon dans la table de passage INE-IDENT transmise à la DEPP (OUI)?" ;
Proc freq data=verif2;
table n;
table prefixIDENT*n/nopercent norow nocol;
run;
Data verif verif2; set _null_;
run;

/* 0.b) Verif unicité des INE dans la table de passage INE-IDENT transmise à DEPP */
Proc sort data=depp.ine_gene2017; by ine;run;
Data verif; set depp.ine_gene2017;
Retain n 0 inel;
If ine=inel then n=n+1;
Else n=1;
Inel=ine;

PrefixIDENT=substr(IDENT,1,2);

IneVide=(INE="");

If substr(INE,11,1)=" " then do; INEbis="INVALIDE";Qualite_INE="INVALIDE";end;
                        *reperage des valeurs trop courtes;
Else if INE*1 ge 0 then do; INEbis="INVALIDE";Qualite_INE="INVALIDE";end;
                        *repérage des valeurs sans lettre;
Else if substr(INE,1,1) in ("A","B","C","D","E","F","G","H","I","J","K","L","M","N","O","P","Q","R","S",

```

```

                "T", "U", "V", "W", "X", "Y", "Z") then do; INEbis="INVALIDE";Qualite_INE="INVALIDE";end;
                *repérage de valeurs commençant par une lettre;
Else if substr(INE,1,5)="99999" then do; INEbis="INVALIDE";Qualite_INE="INVALIDE";end;
                *repérage de valeurs composées de 9 et une lettre;
Else do; INEbis=INE; Qualite_INE="VALIDE ?";end;

run;

Title "Examen table de la qualité des INE dans la table de passage";
Proc freq data=Verif;
tables Qualite_INE PrefixIDENT ;
tables PrefixIDENT*(Qualite_INE IneVide)/nopercent nocol norow;
run;

Data verif2; set verif; by ine; if last.ine;run;
Title "Il y a t'il des INE en doublon dans la table de passage INE-IDENT transmise à la DEPP (PLUTOT NON)?" ;
Proc freq data=verif2; table n;
run;
Title "Liste des 0,3% d'INE en doublons";
Proc print data=verif2 (where=(n ge 2));

Data verif verif2; set _null_;
run;

/*****/
/*****/
/*****/
/***** CONSTRUCTION D'UNE NOUVELLE TABLE DE PASSAGE INE/IDENT */
/***** AYANT UNE OBSERVATION PAR IDENT ET TROIS VARIABLES INE POSSIBLE*/
/***** -> En réalité, travail non indispensable puisque la DEPP */
/***** a injecté les IDENT a la place des INE dans ses tables
/*****/
/*****/
/*****/

*** Table de départ avec variables construites;
Proc sort data=depp.ine_gene2017; by INE; run;
Data ine_gene2017_V2; set depp.ine_gene2017;
Length INEbis $11.;

```

```

PrefixIDENT=substr(IDENT,1,2);
IneVide=(INE="");

If substr(INE,11,1)="" then do; INEbis="INVALIDE";Qualite_INE="INVALIDE";end;
      *reperage des valeurs trop courtes;
Else if INE*1 ge 0 then do; INEbis="INVALIDE";Qualite_INE="INVALIDE";end;
      *repérage des valeurs sans lettre;
Else if substr(INE,1,1) in ("A","B","C","D","E","F","G","H","I","J","K","L","M","N","O","P","Q","R","S",
      "T","U","V","W","X","Y","Z") then do; INEbis="INVALIDE";Qualite_INE="INVALIDE";end;
      *repérage de valeurs commençant par une lettre;
Else if substr(INE,1,5)="99999" then do; INEbis="INVALIDE";Qualite_INE="INVALIDE";end;
      *repérage de valeurs composées de 9 et une lettre;
Else do; INEbis=INE; Qualite_INE="VALIDE ?";end;
run;

*** Identification des 3 lignes possibles pour un même IDENT;
Proc sort data=ine_gene2017_v2; by ident;run;
Data Temp; set ine_gene2017_v2;
Retain n 0 ident1;
If ident=ident1 then n=n+1;
Else n=1;
Ident1=ident;
run;

*** Tables des 1ere observations;
Data Temp1; set Temp; if n=1;
Length INE1 $11. INEbis1 $11.;
Qualite_INE1=Qualite_INE; INE1=INE; INEbis1=INEbis;repondant1=repondant;
Keep IDENT repondant1 PrefixIDENT INE1 INEbis1 Qualite_INE1;
Run;

*** Tables des 2e observation;
Data Temp2; set Temp; if n=2;
Length INE2 $11. INEbis2 $11.;
Qualite_INE2=Qualite_INE; INE2=INE; INEbis2=INEbis;repondant2=repondant;
Keep IDENT repondant2 INE2 INEbis2 Qualite_INE2;
Run;

*** Tables des 3e observation;
Data Temp3; set Temp; if n=3;
Length INE3 $11. INEbis3 $11.;

```

```

Qualite_INE3=Qualite_INE; INE3=INE; INEb3=INEbis;repondant3=repondant;
Keep IDENT repondant3 INE3 INEb3 Qualite_INE3;
Run;

*** Creation de la table avec une ligne par IDENT et les différents INE associés;
Data table_passage; merge Temp1 temp2; by ident;
If repondant2="" then repondant2="9";
If Qualite_INE2="" then Qualite_INE2="ABSENT ";
run;
Data sor.table_passage; merge Table_passage temp3; by ident;
If repondant3="" then repondant3="9";
If Qualite_INE3="" then Qualite_INE3="ABSENT ";
Nbrepondants=(repondant1="1")+(repondant2="1")+(repondant3="1");
NbINE=(repondant1 ne "9")+(repondant2 ne "9")+(repondant3 ne "9");
NbINEValide=(Qualite_INE1="VALIDE ?")+(Qualite_INE2="VALIDE ?")+(Qualite_INE3="VALIDE ?");
run;

Title "Examen de la table de passage créé avec une seule ligne par IDENT";
Proc freq data=sor.table_passage;
tables repondant1 repondant2 repondant3 nbrepondants;
tables Nbrepondants*NbINE Nbrepondants*NbINEValide /nopercnt nocol norow;
run;

Title "Nombre d'INE potentiellement valide selon le préfix de l'IDENT (champ des répondants)";
Proc freq data=sor.table_passage (where=(repondant1="1"));
tables PrefixIdent*NbINEValide /nopercnt nocol norow;
run;

/*****
/*****
/*****
/* 1. ANALYSE EXPLORATOIRE BASIQUE DES TABLES DEPP */
/*****
/*****
/*****

%macro Explo_faere(an);

```

```
Data Explo; set depp.generationfaere&an.;
PrefixIDENT=substr(IDENT,1,2);
Diplome = substr(dipl_spec,1,3);
Exam1 = substr(exam_dipl,1,3);
Exam2 = substr(exam_dipl2,1,3);
run;

***** Attention : variable ine_genere absente en 2007, 2006, 2005;
Title "Exploratoire sur la table FAERE &an. livrée seule";
Proc freq data=Explo;
tables an_sco repondant_generation present_faere&an. ine_generation_fictif source
      flag_doublon type_doublon annee_nais sexe ;
run;

%if %eval(&an.) ge 8 %then %do;
Proc freq data=Explo; tables ine_genere;run;
%end;

Proc freq data=Explo;
tables contrat secteur type_etab3 bourse statut;
tables statut*source /nopercent nocol norow;
tables exam1 pres_dipl result_dipl;
run;

%if %eval(&an.) ge 13 %then %do;
Proc freq data=Explo;
tables motif_sortie exam2 pres_dipl2 result_dipl2;
run;
%end;

***** Calcul du nombre de ligne par IDENT dans la table Faere&an. ;
Proc sort data=depp.generationfaere&an.; by ident;run;
Data Comptage; set depp.generationfaere&an. (where=(an_sco ne ""));

/*
Retain n&an. ident1;
If ident=ident1 then n&an.=n&an.+1;
Else n&an.=1;
Ident1=ident;
*/
Retain n&an. 0 n_sco 0 n_sifa 0 n_flag0 0 n_flag1 0 ident1;
```

```
If ident=ident1 then do;
    n&an.=n&an.+1;
    n_sco=n_sco+(source="SCO");
    n_sifa=n_sifa+(source="SIFA");
    n_flag0=n_flag0+(flag_doublon="0");
    n_flag1=n_flag1+(flag_doublon="1");
end;
Else do;
    n&an.=1;
    n_sco=(source="SCO");
    n_sifa=(source="SIFA");
    n_flag0=(flag_doublon="0");
    n_flag1=(flag_doublon="1");
end;
Ident1=ident;
run;
Proc sort data=comptage; by ident;run;
Data comptage; set comptage; by ident; if last.ident;
run;

Title "Nombre de présences d un IDENT donné dans la table FAERE &an. (champ restreint aux données FAERE non vides)";
Proc freq data=Comptage ;tables n&an. ;
run;

Title "Cas des IDENT présents deux fois dans la table FAERE &an. (champ restreint aux données FAERE non vides)";
Proc freq data=Comptage(where=(n&an.=2)) ;tables n_sco*n_sifa n_flag0*n_flag1/norow nocol;
run;

**** Les Ident avec plusieurs lignes étaient-il en doublon dans la table de passage ?;
Proc sort data=Comptage; by ident;run;
Proc sort data=sor.table_passage;by ident;run;
Data Comptage2; merge comptage (in=a)
                    sor.table_passage (keep=NbINEValide NbINE ident);
by ident; if a;
run;
Title "multiplicité des IDENT dans la table FAERE &an. et dans la table de passage(champ restreint aux données FAERE non vides)";
Proc freq data=Comptage2 ;tables n&an.*(NbINEValide NbINE)/nopercnt norow nocol ;
run;

Data Explo Comptage comptage2; set _null_;
```



```
run;
%mend;
run;

%Explo_faere(16); run;
%Explo_faere(15); run;
%Explo_faere(14); run;
%Explo_faere(13); run;
%Explo_faere(12); run;
%Explo_faere(11); run;
%Explo_faere(10); run;
%Explo_faere(09); run;
%Explo_faere(08); run;
%Explo_faere(07); run;
%Explo_faere(06); run;
%Explo_faere(05); run;

/*****
/*****
/** Construction d'une Table FAERE retravaillée */
/** Après examen des doublons d'IDENT DANS FAERE **/
/** pour avoir une seule ligne par IDENT */
/*****
/*****
%Macro Table_faere(an);

**** Etape 1 : construction des variables;
***** ;
Data Table; set depp.generationfaere&an. (where=(an_sco ne ""));

Aleas=ranuni(1);
*variable utilisée pour choisir de façon aléatoire infra quelle observation;
```

```

*conserver en cas de présences multiples autre que les doublons avec un;
*Flag_doublon=1 et un flag_doublon=0;

PrefixIDENT=substr(IDENT,1,2);

Niv_fms&an.=substr(fortrams7,1,1);
Spec_fms&an.=substr(fortrams7,2,1);
Fin_fms&an.=substr(fortrams7,3,1);
Classe_fms&an.=substr(fortrams7,1,6);

an_sco_&an.=an_sco;
source_&an.=source;
secteur_&an.=secteur;
statut_&an.=statut;
Type_doublon_&an.=Type_doublon;
flag_doublon_&an.=flag_doublon;

%if %eval(&an.) ge 8 %then %do;
Ine_genere_&an.=Ine_genere;
%end;

%if %eval(&an.) ge 11 %then %do;
motif_sortie_&an.=motif_sortie;
date_sortie_&an.=date_sortie;
%end;

*premier examen enregistré; /* Attention, Exam_Dipl peut être différents de Dipl_Spec */
Exam1_&an. = substr(exam_dipl,1,3); If Exam1_&an.="" then Exam1_&an.="ZZZ";
If Exam1_&an. ne "ZZZ" then do;
If substr(exam_dipl,1,3) in ("670") then diplome1_&an.="Brevet ";
else If substr(exam_dipl,1,3) in ("470") then diplome1_&an.="Bac Gén. ";
else If substr(exam_dipl,1,3) in ("430") then diplome1_&an.="Bac techno";
else If substr(exam_dipl,1,3) in ("400") then diplome1_&an.="Bac pro. ";
else If substr(exam_dipl,1,3) in ("500") then diplome1_&an.="CAP/BEP ";
else If substr(exam_dipl,1,3) in ("510") then diplome1_&an.="CAP/BEP ";
else diplome1_&an.="Autre dipl";
end;

If result_dipl in ("11","12","13","14","21","22") then result1_&an.=result_dipl; *resultats des presents;
Else if pres_dipl=1 then result1_&an.="30";*present mais resultats inconnus;
Else if pres_dipl=0 then result1_&an.="00";* inscrit, mais absent aux examens;
Else result1_&an.="ZZ";*non concerné;

```

```

%if %eval(&an.) ge 13 %then %do;
*2e examen enregistré (cas rare);
Exam2_&an. = substr(exam_dipl2,1,3); If Exam2_&an.="" then Exam2_&an.="ZZZ";

If result_dipl2 in ("11","12","13","14","21","22") then result2_&an.=result_dipl2; *resultats des presents;
Else if pres_dipl2=1 then result2_&an.="30";*present mais resultats inconnus;
Else if pres_dipl2=0 then result2_&an.="00";* inscrit, mais absent aux examens;
Else result2_&an.="ZZ";*non concerné;
%end;

***** Classes à partir de FORTRAMS7 (transverse SIFA-SYSCA);
If source="SCO " then           classe&an.="99-Autres SCO           ";
If source="SIFA" then          classe&an.="99-Autres SIFA          ";

If substr(fortrams7,1,1)= "0" then      classe&an.="00-Pré élémentaire  ";
If substr(fortrams7,1,1)= "1" then      classe&an.="01-Primaire      ";

If substr(fortrams7,1,1)= "2" then      classe&an.="07-Niv. Coll-autres  ";
If substr(fortrams7,1,6)= "211114" then classe&an.="02-6e           ";
If substr(fortrams7,1,6)= "211124" then classe&an.="03-5e           ";
If substr(fortrams7,1,6)= "211144" then classe&an.="04-4e générale  ";
If substr(fortrams7,1,6)= "211154" then classe&an.="05-3e générale  ";
If substr(fortrams7,1,4)= "2112" then   classe&an.="06-SES/SEGPA    ";

If substr(fortrams7,1,1)= "3" then      classe&an.="13-Niv. Lycée-autres ";
If substr(fortrams7,1,4)= "3110" then   classe&an.="08-2de générale  ";
If substr(fortrams7,1,4)= "3111" then   classe&an.="08-2de générale  ";
If substr(fortrams7,1,4)= "3112" then   classe&an.="09-1re générale  ";
If substr(fortrams7,1,4)= "3212" then   classe&an.="10-1re techno    ";
If substr(fortrams7,1,4)= "3113" then   classe&an.="11-Terminale générale";
If substr(fortrams7,1,4)= "3213" then   classe&an.="12-Terminale techno ";

If fortrams7="3311011" then             classe&an.="15-CAP TERM      "; *CAP1/1;
If fortrams7="3311021" then             classe&an.="14-CAP          "; *CAP1/2;
If fortrams7="3311022" then             classe&an.="15-CAP TERM      "; *CAP2/2;
If fortrams7="3311031" then             classe&an.="14-CAP          "; *CAP1/3;
If fortrams7="3311032" then             classe&an.="14-CAP          "; *CAP2/3;
If fortrams7="3311033" then             classe&an.="15-CAP TERM      "; *CAP3/3;

If fortrams7="3312011" then             classe&an.="21-BEP TERM      "; *BEP1/1;
If fortrams7="3312021" then             classe&an.="20-BEP          "; *BEP1/2;

```

```

If fortrams7="3312022" then               classe&an.="21-BEP TERM           "; *BEP2/2;

If fortrams7 in ("3313011","3313111") then classe&an.="24-Bac Pro TERM       "; *Bac Pro 1/1;
If fortrams7 in ("3313021","3313121") then classe&an.="23-Bac Pro           "; *Bac Pro 1/2;
If fortrams7 in ("3313022","3313122") then classe&an.="24-Bac Pro TERM       "; *Bac Pro 2/2;
If fortrams7 in ("3313031","3313131") then classe&an.="23-Bac Pro           "; *Bac Pro 1/3;
If fortrams7 in ("3313032","3313132") then classe&an.="23-Bac Pro           "; *Bac Pro 2/3;
If fortrams7 in ("3313033","3313133") then classe&an.="24-Bac Pro TERM       "; *Bac Pro 3/3;

If substr(fortrams7,1,4)="3314" then       classe&an.="29-Mention comple.   ";

If fortrams7="3315021" then classe&an.="30-Brevet Pro           "; * Brevet Pro 1/2;
If fortrams7="3315022" then classe&an.="31-Brevet Pro TERM      "; * Brevet Pro 2/2;
If fortrams7="3315031" then classe&an.="30-Brevet Pro           "; * Brevet Pro 1/3;
If fortrams7="3315032" then classe&an.="30-Brevet Pro           "; * Brevet Pro 2/3;
If fortrams7="3315033" then classe&an.="31-Brevet Pro TERM      "; * Brevet Pro 3/3;

If substr(fortrams7,1,1)="4" then          classe&an.="35-Fortrams1=4           ";

If substr(fortrams7,1,1)="5" then          classe&an.="46-Autre Fortrams1=5 ";
If fortrams7="5141121" then classe&an.="36-CPGE 1/2           ";
If fortrams7="5141122" then classe&an.="37-CPGE 2/2           ";

If fortrams7 in ("5411011","5411111","5411211") then classe&an.="39-BTS TERM           ";* BTS 1/1;
If fortrams7 in ("5411021","5411121","5411221") then classe&an.="38-BTS           ";* BTS 2/1;
If fortrams7 in ("5411022","5411122","5411222") then classe&an.="39-BTS TERM           ";* BTS 2/2;
If fortrams7 in ("5411031","5411131","5411231") then classe&an.="38-BTS           ";* BTS 1/3;
If fortrams7 in ("5411032","5411132","5411232") then classe&an.="38-BTS           ";* BTS 2/3;
If fortrams7 in ("5411033","5411133","5411233") then classe&an.="39-BTS TERM           ";* BTS 3/3;

If fortrams7="5412121" then classe&an.="44-DUT 1/2           ";
If fortrams7="5412122" then classe&an.="45-DUT 2/2           ";

If substr(fortrams7,1,1)="6" then          classe&an.="58-Autre Fortrams1=6 ";
If substr(fortrams7,1,6)="641121" then classe&an.="47-Licence Pro           ";

If fortrams7="6417121" then classe&an.="48-Master Pro 1/2      ";
If fortrams7="6417122" then classe&an.="49-Master Pro 2/2      ";

If substr(fortrams7,1,4)="6421" then classe&an.="53-Ingenieur DIV           ";
If fortrams7="6421131" then classe&an.="50-Ingenieur           ";*Ingenieur 1/3;
If fortrams7="6421132" then classe&an.="50-Ingenieur           ";*Ingenieur 2/3;

```

```

If fortrams7="6421133" then classe&an.="52-Ingenieur TERM      ";*Ingenieur 3/3;

If substr(fortrams7,1,4)="6422" then classe&an.="57-Commerce  DIV      ";
If fortrams7="6422131" then classe&an.="54-Commerce          ";*Commerce 1/3;
If fortrams7="6422132" then classe&an.="54-Commerce          ";*Commerce 2/3;
If fortrams7="6422133" then classe&an.="56-Commerce TERM      ";*Commerce 3/3;
/**
if substr(Mefst11,1,1)= "4"      then classe&an.="Sys42-Raccrochage      ";
**/

CS&an.=PCSr; If CS&an.=" " then CS&an.="ZZ";
*PCSr est la PCS déclaré dans SIFA et, en 2008, dans SYSCA;
* sinon c est la PCS du père qui est retenue quand elle est dispo dans SYSCA;

If Natio="100" then Natio&an.="France";
Else if substr(Natio,1,1)="1" then Natio&an.="Europe ";
Else if substr(Natio,1,1)="2" then Natio&an.="Asie  ";* inclus la Turquie;
Else if substr(Natio,1,1)="3" then Natio&an.="Africa ";
Else if substr(Natio,1,1)="4" then Natio&an.="America";
Else if substr(Natio,1,1)="5" then Natio&an.="Oceanie";
Else if substr(Natio,1,1)="9" then Natio&an.="Nd+apat";
Else Natio&an.="ZZZZZZZ";
* peu de détail dans SIFA;

Boursier&an.=bourse;if Boursier&an.=" " then Boursier&an.="Z";
*absent de SIFA;

Dpt&an.=substr(com_resid,1,2);

AcadResid&an.="ZZZZZZZZZ";
If dpt&an. in ("25","39","70","90")      then AcadResid&an.="Besançon";
If dpt&an. in ("24","33","40","47","64") then AcadResid&an.="Bordeaux";
If dpt&an. in ("14","50","61")           then AcadResid&an.="Caen      ";
If dpt&an. in ("03","15","43","63")      then AcadResid&an.="Clermont";
If dpt&an. in ("2A","2B","20")           then AcadResid&an.="Corte   ";
If dpt&an. in ("77","93","94")           then AcadResid&an.="Créteil  ";
If dpt&an. in ("21","58","71","89")      then AcadResid&an.="Dijon    ";
If dpt&an. in ("07","26","38","73","74") then AcadResid&an.="Grenoble";
If dpt&an. in ("97")                     then AcadResid&an.="DOM      ";
If dpt&an. in ("59","62")                 then AcadResid&an.="Lille    ";
If dpt&an. in ("19","23","87")           then AcadResid&an.="Limoges  ";

```

```

If dpt&an. in ("01","42","69") then AcadResid&an.="Lyon ";
If dpt&an. in ("11","30","34","48","66") then AcadResid&an.="Montpell";
If dpt&an. in ("54","55","57","88") then AcadResid&an.="NancMetz";
If dpt&an. in ("44","49","53","72","85") then AcadResid&an.="Nantes ";
If dpt&an. in ("06","83") then AcadResid&an.="Nice ";
If dpt&an. in ("16","17","79","86") then AcadResid&an.="Poitiers";
If dpt&an. in ("08","10","51","52") then AcadResid&an.="Reims ";
If dpt&an. in ("22","29","35","56") then AcadResid&an.="Rennes ";
If dpt&an. in ("27","76") then AcadResid&an.="Rouen ";
If dpt&an. in ("67","68") then AcadResid&an.="Strasbou";
If dpt&an. in ("09","12","31","32","46","65","81","82")
then AcadResid&an.="Toulouse";

If dpt&an. in ("78","91","92","95") then AcadResid&an.="Versaill";
If dpt&an. in ("04","05","13","84") then AcadResid&an.="AixMarse";
If dpt&an. in ("02","60","80") then AcadResid&an.="Amiens ";
If dpt&an. in ("18","28","36","37","41") then AcadResid&an.="OrléTour";

%if %eval(&an.) ge 13 %then %do;
Keep an_sco_&an. classe&an. IDENT aleas
secteur_&an. statut_&an. CS&an. Natio&an. Dpt&an. Boursier&an.
AcadResid&an.
Type_doublon_&an. flag_doublon_&an. source_&an. sexe annee_nais
Exam1_&an. Exam2_&an. Result1_&an. Result2_&an.
Diplome1_&an. Pres_dipl result_dipl Pres_dipl2 result_dipl2
Niv_fms&an. Spec_fms&an. Fin_fms&an. Classe_fms&an.
Ine_genere_&an. motif_sortie_&an. date_sortie_&an.;
%end;

* Pour 2012 et avant, un seul examen renseigné;
%if %eval(&an.) ge 11 and %eval(&an.) le 12 %then %do;
Keep an_sco_&an. classe&an. IDENT aleas
secteur_&an. statut_&an. CS&an. Natio&an. Dpt&an. Boursier&an.
AcadResid&an.
Type_doublon_&an. flag_doublon_&an. source_&an. sexe annee_nais
Exam1_&an. Result1_&an.
Diplome1_&an. Pres_dipl result_dipl
Niv_fms&an. Spec_fms&an. Fin_fms&an. Classe_fms&an.
Ine_genere_&an. motif_sortie_&an. date_sortie_&an.;
%end;

* Pour 2010 et avant, pas de date et motif de sortie;
%if %eval(&an.) ge 8 and %eval(&an.) le 10 %then %do;

```

```
Keep an_sco_&an. classe&an. IDENT aleas
    secteur_&an. statut_&an. CS&an. Natio&an. Dpt&an. Boursier&an.
    AcadResid&an.
    Type_doublon_&an. flag_doublon_&an. source_&an. sexe annee_nais
    Exam1_&an. Result1_&an.
    Diplome1_&an. Pres_dipl result_dipl
    Niv_fms&an. Spec_fms&an. Fin_fms&an. Classe_fms&an.
    Ine_genere_&an.;
%end;

* Pour 2007 et avant, pas de variable INE_GENERE;
%if %eval(&an.) le 7 %then %do;
Keep an_sco_&an. classe&an. IDENT aleas
    secteur_&an. statut_&an. CS&an. Natio&an. Dpt&an. Boursier&an.
    AcadResid&an.
    Type_doublon_&an. flag_doublon_&an. source_&an. sexe annee_nais
    Exam1_&an. Result1_&an.
    Diplome1_&an. Pres_dipl result_dipl
    Niv_fms&an. Spec_fms&an. Fin_fms&an. Classe_fms&an.;
%end;
run;

Title "Avec FAERE&an. -Verification des variables construites dans TABLE avec FORTRAMS_7";
Proc freq data=Table;
tables classe&an.*(Niv_fms&an. Spec_fms&an. Fin_fms&an.)/nopercent nocol norow;
run;
Proc sort data=Table; by classe&an.;run;
Proc freq data=Table;
tables Classe_fms&an. ; by classe&an.;
run;

Title "Avec FAERE&an. -Verification des variables construites dans TABLE";
Proc freq data=Table;
tables exam1_&an. result1_&an. Pres_dipl result_dipl;
tables (exam1_&an. classe&an.)*diplome1_&an./nopercent norow nocol;
tables AcadResid&an.;
run;
%if %eval(&an.) ge 13 %then %do;
Proc freq data=Table;
tables exam2_&an. result2_&an. Pres_dipl2 result_dipl2;
run;
%end;
```

```

**** Etape 2 : Construction de la table "FAERE" avec une seule ligne par IDENT
****          et récupération des variables importantes des différentes lignes;
****          Pour cela, on trie la table d'entrée par IDENT en mettant, pour ;
****          un IDENT donné, le ou les flag_doublon=1 en premier;
***** ;

```

```
Proc sort data=Table; by ident descending flag_doublon_&an. aleas; run;
```

```
***** pour les tables 2013 et suivantes;
```

```
%if %eval(&an.) ge 13 %then %do;
```

```
Data Table; set Table (drop=Pres_dipl result_dipl Pres_dipl2 result_dipl2);
```

```
Retain n&an. 0 ident1;
```

```
If ident=ident1 then do;
```

```
  n&an.=n&an.+1;
```

```
  If n&an.=2 then do; source_L2_&an.=source_&an.; Classe&an._L2 = classe&an.;
```

```
    diplome_L2_&an.=Diplome1_&an.; statut_L2_&an.=Statut_&an.;
```

```
    Ine_genere_L2_&an.=Ine_genere_&an.;
```

```
    motif_sortie_L2_&an.=motif_sortie_&an.; date_sortie_L2_&an.=date_sortie_&an.;
```

```
    Type_doublon_L2_&an.=type_doublon_&an.; Flag_doublon_L2_&an.=flag_doublon_&an.;
```

```
      Exam1_L2_&an.=Exam1_&an.; Result1_L2_&an.=Result1_&an.;
```

```
      Exam2_L2_&an.=Exam2_&an.; Result2_L2_&an.=Result2_&an.;
```

```
    end;
```

```
  If n&an.=3 then do; source_L3_&an.=source_&an.; Classe&an._L3 = classe&an.;
```

```
    diplome_L3_&an.=Diplome1_&an.; statut_L3_&an.=Statut_&an.;
```

```
    Ine_genere_L3_&an.=Ine_genere_&an.;
```

```
    motif_sortie_L3_&an.=motif_sortie_&an.; date_sortie_L3_&an.=date_sortie_&an.;
```

```
    Type_doublon_L3_&an.=type_doublon_&an.; Flag_doublon_L3_&an.=flag_doublon_&an.;
```

```
      Exam1_L3_&an.=Exam1_&an.; Result1_L3_&an.=Result1_&an.;
```

```
      Exam2_L3_&an.=Exam2_&an.; Result2_L3_&an.=Result2_&an.;
```

```
    end;
```

```
  end;
```

```
Else n&an.=1;
```

```
Ident1=ident;
```

```
run;
```

```
%end;
```



```

***** pour années antérieures;
%if %eval(&an.) le 12 %then %do;
Data Table; set Table (drop=Pres_dipl result_dipl);
Retain n&an. 0 ident1;

If ident=ident1 then do;
    n&an.=n&an.+1;
    If n&an.=2 then do; source_L2_&an.=source_&an.; Classe&an._L2 = classe&an.;
        diplome_L2_&an.=Diplome1_&an.; statut_L2_&an.=Statut_&an.;
        Type_doublon_L2_&an.=type_doublon_&an.; Flag_doublon_L2_&an.=flag_doublon_&an.;
        Exam1_L2_&an.=Exam1_&an.; Result1_L2_&an.=Result1_&an.;
    end;
    If n&an.=3 then do; source_L3_&an.=source_&an.; Classe&an._L3 = classe&an.;
        diplome_L3_&an.=Diplome1_&an.; statut_L3_&an.=Statut_&an.;
        Type_doublon_L3_&an.=type_doublon_&an.; Flag_doublon_L3_&an.=flag_doublon_&an.;
        Exam1_L3_&an.=Exam1_&an.; Result1_L3_&an.=Result1_&an.;
    end;

end;
Else n&an.=1;
Ident1=ident;
run;
%end;

*****;

*table des lère observations liées à un IDENT donné;
Data Table_part1; set table ; if n&an.=1;
Drop n&an.;
run;

```

```

*table des 2e observations liées à un IDENT donné;
Data Table_part2; set table ; if n&an.=2;

%if %eval(&an.) ge 13 %then %do;
Drop n&an. diplome1_&an. statut_&an. Ine_genere_&an. motif_sortie_&an. date_sortie_&an.
type_doublon_&an. flag_doublon_&an.
CS&an. Natio&an. Dpt&an. Boursier&an. AcadResid&an.
exam1_&an. result1_&an. exam2_&an. result2_&an. source_&an. classe&an.;
%end;

%if %eval(&an.) le 12 %then %do;
Drop n&an. diplome1_&an. statut_&an.
type_doublon_&an. flag_doublon_&an.
CS&an. Natio&an. Dpt&an. Boursier&an. AcadResid&an.
exam1_&an. result1_&an. source_&an. classe&an.;
%end;
run;

*table des 3e observations liées à un IDENT donné;
Data Table_part3; set table ; if n&an.=3;
%if %eval(&an.) ge 13 %then %do;
Drop n&an. diplome1_&an. statut_&an. Ine_genere_&an. motif_sortie_&an. date_sortie_&an.
type_doublon_&an. flag_doublon_&an.
CS&an. Natio&an. Dpt&an. Boursier&an. AcadResid&an.
exam1_&an. result1_&an. exam2_&an. result2_&an. source_&an. classe&an.;
%end;

%if %eval(&an.) le 12 %then %do;
Drop n&an. diplome1_&an. statut_&an.
CS&an. Natio&an. Dpt&an. Boursier&an. AcadResid&an.
type_doublon_&an. flag_doublon_&an.
exam1_&an. result1_&an. source_&an. classe&an.;
%end;

*Constitution de la table finale;
Data sor.Table&an.; Merge Table_part1 Table_part2; by ident; run;
Data sor.Table&an.; Merge sor.Table&an. Table_part3; by ident; run;

Title "Table&an. -Verification de la table construite avec une observation par IDENT";
Proc freq data=sor.Table&an.;
table source_&an. source_L2_&an. source_L3_&an.;
table flag_doublon_&an. flag_doublon_L2_&an. flag_doublon_L3_&an.;

```

```
Table flag_doublon_&an.*(flag_doublon_L2_&an. flag_doublon_L3_&an.)/nopercent nocol norow;
Table flag_doublon_L2_&an.*flag_doublon_L3_&an./nopercent nocol norow;
table Type_doublon_&an. type_doublon_L2_&an. type_doublon_L3_&an.;
run;
```

```
%if %eval(&an.) ge 13 %then %do;
Proc freq data=sor.Table&an.;
table ine_genere_&an. ine_genere_L2_&an. ine_genere_L3_&an.;
run;
%end;
```

```
proc contents data=sor.Table&an.;run;
```

```
Data Table Table_part1 Table_part2; set _null_; run;
```

```
%mend;
run;
```

```
%Table_faere(16);run;
%Table_faere(15);run;
%Table_faere(14);run;
%Table_faere(13);run;
%Table_faere(12);run;
%Table_faere(11);run;
%Table_faere(10);run;
%Table_faere(09);run;
%Table_faere(08);run;
%Table_faere(07);run;
%Table_faere(06);run;
%Table_faere(05);run;
```

```
/******/
/******/
***** 2. Creation de tables annuelles synthetiques SOR.APP&anTable ;
***** en appariant Generation avec les TABLE&AN créées au-dessus
***** DONC RESTREINT AUX IDENT REpondant
/******/
/******/
```

```

***** enrichissement de la table GENERATION par la table de passage INE_IDENT créée supra;
Data Generation; set gene.g17individus; run;
Proc sort data=Generation; by ident; run;
Proc sort data=sor.table_passage ; by ident;

Data Generation ; merge Generation (in=a) sor.table_passage ;
by ident; if a;
PrefixIDENT=substr(IDENT,1,2);

* 2A-qualité de récupération INE dans le champs d'appartenance pour base sondage;
If PrefixIDENT in ("SI","SY","UN") then champBase="INE";
else if PrefixIDENT in ("SP","CI","CP") then champBase="NON";
else if PrefixIDENT in ("AG","CU","G1","G3","G4") then champBase="BOF";

* PHD : deux premières positions de PHD_CEREQ;
* PHD_CEREQ (niveau du plus haut diplôme dans la nomenclature Céreq) ;

phd=substr(phd_cereq,1,2); if phd="" then phd="ZZ";
nivsor=substr(nivsor_cereq,1,2); if nivsor="" then nivsor="ZZ";

** Codification du plus haut diplôme pour le QEEF (message Manon 17 janvier 2021);
If phd_cereq="01" then phd_qeef="01-Non diplômé ";
If phd_cereq="02I" then phd_qeef="02-CAP/BEP Indus. ";
If phd_cereq="02T" then phd_qeef="03-CAP/BEP Tert. ";
If phd_cereq="03I" then phd_qeef="04-Bac Pro/techno Indus. ";
If phd_cereq="04I" then phd_qeef="04-Bac Pro/techno Indus. ";
If phd_cereq="05I" then phd_qeef="04-Bac Pro/techno Indus. ";
If phd_cereq="03T" then phd_qeef="05-Bac Pro/techno Tert. ";
If phd_cereq="04T" then phd_qeef="05-Bac Pro/techno Tert. ";
If phd_cereq="05T" then phd_qeef="05-Bac Pro/techno Tert. ";
If phd_cereq="06" then phd_qeef="06-Bac général ";
If phd_cereq="07I" then phd_qeef="07-Bac+2 Indus. ";
If phd_cereq="08I" then phd_qeef="07-Bac+2 Indus. ";
If phd_cereq="07T" then phd_qeef="08-Bac+2 Tert. ";
If phd_cereq="08T" then phd_qeef="08-Bac+2 Tert. ";
If phd_cereq="09" then phd_qeef="09-Bac+2à4 Santé Social ";
If phd_cereq="10L" then phd_qeef="10-Bac+3à4/licence pro. Litt./tert.";
If phd_cereq="11L" then phd_qeef="10-Bac+3à4/licence pro. Litt./tert.";
If phd_cereq="12L" then phd_qeef="10-Bac+3à4/licence pro. Litt./tert.";
If phd_cereq="10M" then phd_qeef="11-Bac+3à4/licence pro. Sci./indus.";
If phd_cereq="11M" then phd_qeef="11-Bac+3à4/licence pro. Sci./indus.";

```

```

If phd_cereq="12M" then      phd_qeef="11-Bac+3à4/licence pro. Sci./indus.";
If phd_cereq="13L" then      phd_qeef="12-Bac+5 Litt.                ";
If phd_cereq="14L" then      phd_qeef="12-Bac+5 Litt.                ";
If phd_cereq="13M" then      phd_qeef="13-Bac+5 Sci.                  ";
If phd_cereq="14M" then      phd_qeef="13-Bac+5 Sci.                  ";
If phd_cereq="15" then       phd_qeef="14-Ecole de commerce          ";
If phd_cereq="16" then       phd_qeef="15-Ecole d'ingénieur          ";
If phd_cereq="17" then       phd_qeef="16-Doctorat                   ";
If phd_cereq="18L" then      phd_qeef="16-Doctorat                   ";
If phd_cereq="18M" then      phd_qeef="16-Doctorat                   ";
If phd_qeef="" then         phd_qeef="ZZ                               ";

** Codification du plus haut diplôme plus agrégée pour le QEEF (message Manon 17 janvier 2021);
If substr(phd_qeef,1,2) = "01" then      phd_qeef2="01-Non diplômés      ";
If substr(phd_qeef,1,2) in ("06","03") then      phd_qeef2="02-CAP-BEP-MC-Divers Niv.3";
If phd_cereq in ("03I","04I","03T","04T") then      phd_qeef2="03-Bac Pro-BT-Divers Niv.4";
If phd_cereq in ("05I","05T") then          phd_qeef2="04-Bac technologique      ";
If substr(phd_qeef,1,2) = "06" then          phd_qeef2="05-Bac général          ";
If substr(phd_qeef,1,2) in ("07","08") then      phd_qeef2="06-Bac+2 sauf Santé/Social";
If substr(phd_qeef,1,2) = "09" then          phd_qeef2="07-Bac+2 à +4 Santé/Social";
If phd_cereq in ("10L","10M") then          phd_qeef2="08-Licence professionnelle";
If phd_cereq in ("11L","12L","11M","12M") then      phd_qeef2="09-Autre Bac+3 et Bac+4      ";
If substr(phd_qeef,1,2) in ("12","13") then      phd_qeef2="10-Bac+5 sauf écoles      ";
If substr(phd_qeef,1,2) in ("14","15") then      phd_qeef2="11-Ecoles (Ingé,commerce) ";
If substr(phd_qeef,1,2) = "16" then          phd_qeef2="12-Doctorat          ";
If phd_qeef2="" then          phd_qeef2="ZZ          ";

run;

Title "2A-lien entre le niveau de sortie et le prefixe de l'IDENT (donc la qualité de l'INE)";
Proc freq data=Generation;
tables NIVSOR*(prefixIDENT ChampBase qualite_ine1 qualite_ine2 qualite_ine3)/nopercent norow nocol;
tables prefixIDENT*ChampBase/nopercent norow nocol;
run;
Proc freq data=Generation;
tables NIVSOR*(prefixIDENT ChampBase)/nopercent nofreq nocol;
weight ponddef;
run;

Title "2A-lien entre le plus haut diplôme le prefixe de l'IDENT (donc la qualité de l'INE)";
Proc freq data=Generation;
tables phd_qeef2*(prefixIDENT ChampBase qualite_ine1 qualite_ine2 qualite_ine3)/nopercent norow nocol;

```

```

run;
Proc freq data=Generation;
tables phd_qeef2*(prefixIDENT ChampBase qualite_ine1 qualite_ine2 qualite_ine3)/nopercent nofreq nocol;
weight pondef;
run;

Title "2A-interruption des études avant la sortie";
Title2 "1=Oui // 2=Non";
Proc freq data=Generation;
tables (phd_qeef2 PrefixIDENT)*QF200/nopercent nocol nofreq;
weight pondef;
run;
Title2 "1=moins de 12 mois //2= 13-16 mois // 3 =17 mois ou plus pour raison de santé";
Proc freq data=Generation;
tables (phd_qeef2 PrefixIDENT)*QF210_MAJ/nopercent nocol nofreq;
weight pondef;
run;
Title2 "1=Oui // 2=Non (ATTENTION : CHAMP RESTREINT aux sortants du supérieur";
Proc freq data=Generation;
tables (phd_qeef2 PrefixIDENT)*QF230/nopercent nocol nofreq;
weight pondef;
run;

/*****
/* Attention : la macro ci-dessous nécessite d'avoir dans la WORK */
/* la table Generation créée au dessus */
*****/

%macro App_FaereGene(an);

**** Appariement de la table issue de FAERE créée supra (une ligne par IDENT)avec Génération ;
Proc sort data=sor.Table&an.;by ident;run;
Proc sort data=Generation;by ident;run;

Data Exploratoire; merge sor.Table&an (rename=(annee_nais=annee_nais&an. sexe=sexe&an.))
                        Generation;
                        by ident;

If an_sco_&an.=" " then an_sco_&an.="ZZZZ";
present_faere&an. = 1-(an_sco_&an.="ZZZZ");

```

```

Rep2017Diff=(pondef gt 0);

If pondef="" and present_faere&an.=0 then appariement&an.="G0/F0";
else if pondef="" and present_faere&an.=1 then appariement&an.="G0/F1";
else if pondef gt 0 and present_faere&an.=1 then appariement&an.="G1/F1";
else if pondef gt 0 and present_faere&an.=0 then appariement&an.="G1/F0";

Annais_depp=1*annee_nais&an.;
Annais_cereq=2020-AGE20;

if appariement&an.="G1/F1" then do;
    Meme_sexe&an.=(sexe&an.=sexe);
    Meme_Annais&an.=(annais_depp=annais_cereq);
    CompaSexAnnais="S"!!Meme_sexe&an.!!"Annais"!!Meme_Annais&an.;
end;
Else do; Meme_sexe&an.=9;Meme_Annais&an.=9;
end;

***** Sauvegarde en sortie de la table appariée;
Data sor.App&an.; set Exploratoire;

Title "2B-TABLE&an. (1 ligne par IDENT) appariée avec Génération - Exploratoire";
Proc freq data=Exploratoire;
tables appariement&an.*(Rep2017Diff present_faere&an.)
/nopercent nocol norow;
tables (nivsor phd_qeef2 prefixIdent)*appariement&an./norow nopercent nocol;
tables Meme_sexe&an. Meme_Annais&an. CompaSexAnnais classe&an.;
tables nivsor*classe&an./norow nopercent nocol;
run;
%mend;
run;

%App_FaereGene(16);run;
%App_FaereGene(15);run;
%App_FaereGene(14);run;
%App_FaereGene(13);run;
%App_FaereGene(12);run;
%App_FaereGene(11);run;
%App_FaereGene(10);run;
%App_FaereGene(09);run;
%App_FaereGene(08);run;

```

```

%App_FaereGene (07) ;run;
%App_FaereGene (06) ;run;
%App_FaereGene (05) ;run;

/*****
/*****
***** 3 à 5. Qui est présent combien de fois dans FAERE ?
***** et creation de table agrégée SOR.HISTORIQUE;
/*****
/*****

/***** ATTEntion : supprimer les variables posant des problèmes de format
dans la phase d'appariement
de tous les fichiers annuel dans la phase 3 infra
Drop motif_sortie_&an. motif_sortie_L2_&an. motif_sortie_L3_&an.
date_sortie_&an. date_sortie_L2_&an. date_sortie_L3_&an.
ine_genere_&an. ine_genere_L1_&an. ine_genere_L3_&an.;
run;
*****/
Proc sort data=sor.app16;by ident;run;
Proc sort data=sor.app15;by ident;run;
Proc sort data=sor.app14;by ident;run;
Proc sort data=sor.app13;by ident;run;
Proc sort data=sor.app12;by ident;run;
Proc sort data=sor.app11;by ident;run;
Proc sort data=sor.app10;by ident;run;
Proc sort data=sor.app09;by ident;run;
Proc sort data=sor.app08;by ident;run;
Proc sort data=sor.app07;by ident;run;
Proc sort data=sor.app06;by ident;run;
Proc sort data=sor.app05;by ident;run;

/****
Data sor.Historique19oct; Merge sor.app16 sor.app15 sor.app14 sor.app13 sor.app12 sor.app11
                             sor.app10 sor.app09 sor.app08 sor.app07 sor.app06 sor.app05;
by ident;
run;
****/

Data historique; set sor.Historique19oct;
PrefixIdent=substr(Ident,1,2);

```



```

an16=(appariement16="G1/F1");* repérage des répondants appariés avec FAERE 2016 ;
pres16=2016*an16; *variable de présence en 2016 pour calculer infra la 1ère et dernière année de présence;
if pres16=0 then pres16=.;
an15=(appariement15="G1/F1"); pres15=2015*an15; if pres15=0 then pres15=.;* idem 2015;
an14=(appariement14="G1/F1"); pres14=2014*an14; if pres14=0 then pres14=.;* Etc...;
an13=(appariement13="G1/F1"); pres13=2013*an13; if pres13=0 then pres13=.;
an12=(appariement12="G1/F1"); pres12=2012*an12; if pres12=0 then pres12=.;
an11=(appariement11="G1/F1"); pres11=2011*an11; if pres11=0 then pres11=.;
an10=(appariement10="G1/F1"); pres10=2010*an10; if pres10=0 then pres10=.;
an09=(appariement09="G1/F1"); pres09=2009*an09; if pres09=0 then pres09=.;
an08=(appariement08="G1/F1"); pres08=2008*an08; if pres08=0 then pres08=.;
an07=(appariement07="G1/F1"); pres07=2007*an07; if pres07=0 then pres07=.;
an06=(appariement06="G1/F1"); pres06=2006*an06; if pres06=0 then pres06=.;
an05=(appariement05="G1/F1"); pres05=2005*an05; if pres05=0 then pres05=.;

*** repérage de la première et dernière année de présence;
lastyear=max(pres16,pres15,pres14,pres13,pres12,pres11,pres10,pres09,pres08,pres07,pres06,pres05);
firstyear=min(pres16,pres15,pres14,pres13,pres12,pres11,pres10,pres09,pres08,pres07,pres06,pres05);

*** Nombre d'année de présence, dans le public et le privé;
NbPublic=(secteur_16="1")+(secteur_15="1")+(secteur_14="1")+(secteur_13="1")+
(secteur_12="1")+(secteur_11="1")+(secteur_10="1")+(secteur_09="1")+
(secteur_08="1")+(secteur_07="1")+(secteur_06="1")+(secteur_05="1");

NbPrive=(secteur_16="2")+(secteur_15="2")+(secteur_14="2")+(secteur_13="2")+
(secteur_12="2")+(secteur_11="2")+(secteur_10="2")+(secteur_09="2")+
(secteur_08="2")+(secteur_07="2")+(secteur_06="2")+(secteur_05="2");

Nb_presence=an16+an15+an14+an13+an12+an11+an10+an09+an08+an07+an06+an05;

*** repéré comme boursier au moins une fois au collège (SIFA n'a pas l'info sur les boursiers);

BourseCol=((Boursier16="O" and substr(classe16,1,2) in ("01","02","03","04","05","06"))
+(Boursier15="O" and substr(classe15,1,2) in ("01","02","03","04","05","06"))
+(Boursier14="O" and substr(classe14,1,2) in ("01","02","03","04","05","06"))
+(Boursier13="O" and substr(classe13,1,2) in ("01","02","03","04","05","06"))
+(Boursier12="O" and substr(classe12,1,2) in ("01","02","03","04","05","06"))
+(Boursier11="O" and substr(classe11,1,2) in ("01","02","03","04","05","06"))
+(Boursier10="O" and substr(classe10,1,2) in ("01","02","03","04","05","06"))
+(Boursier09="O" and substr(classe09,1,2) in ("01","02","03","04","05","06"))
+(Boursier08="O" and substr(classe08,1,2) in ("01","02","03","04","05","06"))

```

```

+ (Boursier07="O" and substr(classe07,1,2) in ("01","02","03","04","05","06"))
+ (Boursier06="O" and substr(classe06,1,2) in ("01","02","03","04","05","06"))
+ (Boursier05="O" and substr(classe05,1,2) in ("01","02","03","04","05","06")) ge 1);

*indicatrice de "repéré au moins une fois dans une classe concernée";
PresCol=((substr(classe16,1,2) in ("01","02","03","04","05","06"))
+ (substr(classe15,1,2) in ("01","02","03","04","05","06"))
+ (substr(classe14,1,2) in ("01","02","03","04","05","06"))
+ (substr(classe13,1,2) in ("01","02","03","04","05","06"))
+ (substr(classe12,1,2) in ("01","02","03","04","05","06"))
+ (substr(classe11,1,2) in ("01","02","03","04","05","06"))
+ (substr(classe10,1,2) in ("01","02","03","04","05","06"))
+ (substr(classe09,1,2) in ("01","02","03","04","05","06"))
+ (substr(classe08,1,2) in ("01","02","03","04","05","06"))
+ (substr(classe07,1,2) in ("01","02","03","04","05","06"))
+ (substr(classe06,1,2) in ("01","02","03","04","05","06"))
+ (substr(classe05,1,2) in ("01","02","03","04","05","06")) ge 1);

*** repéré comme boursier au moins une fois au lycée (SIFA n'a pas l'info sur les boursiers);

BourseLyc=((Boursier16="O" and substr(classe16,1,2) in ("08","09","10","11"))
+ (Boursier15="O" and substr(classe15,1,2) in ("08","09","10","11"))
+ (Boursier14="O" and substr(classe14,1,2) in ("08","09","10","11"))
+ (Boursier13="O" and substr(classe13,1,2) in ("08","09","10","11"))
+ (Boursier12="O" and substr(classe12,1,2) in ("08","09","10","11"))
+ (Boursier11="O" and substr(classe11,1,2) in ("08","09","10","11"))
+ (Boursier10="O" and substr(classe10,1,2) in ("08","09","10","11"))
+ (Boursier09="O" and substr(classe09,1,2) in ("08","09","10","11"))
+ (Boursier08="O" and substr(classe08,1,2) in ("08","09","10","11"))
+ (Boursier07="O" and substr(classe07,1,2) in ("08","09","10","11"))
+ (Boursier06="O" and substr(classe06,1,2) in ("08","09","10","11"))
+ (Boursier05="O" and substr(classe05,1,2) in ("08","09","10","11")) ge 1);

*indicatrice de "repéré au moins une fois dans une classe concernée";
PresLyc=((substr(classe16,1,2) in ("08","09","10","11"))
+ (substr(classe15,1,2) in ("08","09","10","11"))
+ (substr(classe14,1,2) in ("08","09","10","11"))
+ (substr(classe13,1,2) in ("08","09","10","11"))
+ (substr(classe12,1,2) in ("08","09","10","11"))
+ (substr(classe11,1,2) in ("08","09","10","11"))
+ (substr(classe10,1,2) in ("08","09","10","11"))
+ (substr(classe09,1,2) in ("08","09","10","11"))

```

```

+ (substr(classe08,1,2) in ("08","09","10","11"))
+ (substr(classe07,1,2) in ("08","09","10","11"))
+ (substr(classe06,1,2) in ("08","09","10","11"))
+ (substr(classe05,1,2) in ("08","09","10","11")) ge 1);

*** Nombre d'années d'absence entre la première et dernière année de présence;
nbtrous=(Lastyear-Firstyear)+1-NB_presence;

*** repérage de la première classe connue;
if pres16=firstyear then firstclasse=classe16;
if pres15=firstyear then firstclasse=classe15;
if pres14=firstyear then firstclasse=classe14;
if pres13=firstyear then firstclasse=classe13;
if pres12=firstyear then firstclasse=classe12;
if pres11=firstyear then firstclasse=classe11;
if pres10=firstyear then firstclasse=classe10;
if pres09=firstyear then firstclasse=classe09;
if pres08=firstyear then firstclasse=classe08;
if pres07=firstyear then firstclasse=classe07;
if pres06=firstyear then firstclasse=classe06;
if pres05=firstyear then firstclasse=classe05;

*** repérage de la dernière classe connue;
if pres16=lastyear then lastclasse=classe16;
if pres15=lastyear then lastclasse=classe15;
if pres14=lastyear then lastclasse=classe14;
if pres13=lastyear then lastclasse=classe13;
if pres12=lastyear then lastclasse=classe12;
if pres11=lastyear then lastclasse=classe11;
if pres10=lastyear then lastclasse=classe10;
if pres09=lastyear then lastclasse=classe09;
if pres08=lastyear then lastclasse=classe08;
if pres07=lastyear then lastclasse=classe07;
if pres06=lastyear then lastclasse=classe06;
if pres05=lastyear then lastclasse=classe05;

/**** ATTENTION : COMPTAGE DE CONNAISSANCES DE CLASSE FAIT SELON LES MODALITES DE LA VARIABLE
CLASSE&AN CONSTRUITE SUPRA EN ETAPE 1 DE PARTIE 1, DONC VERIFIER SI ELLE A ETE MODIFIEE
*****/

```

```

NbPrim=(substr(classe16,1,2)="01")+(substr(classe15,1,2)="01")+(substr(classe14,1,2)="01")
+ (substr(classe13,1,2)="01")+(substr(classe12,1,2)="01")+(substr(classe11,1,2)="01")
+ (substr(classe10,1,2)="01")+(substr(classe09,1,2)="01")+(substr(classe08,1,2)="01")
+ (substr(classe07,1,2)="01")+(substr(classe06,1,2)="01")+(substr(classe05,1,2)="01");

Nb6eme=(substr(classe16,1,2)="02")+(substr(classe15,1,2)="02")+(substr(classe14,1,2)="02")
+ (substr(classe13,1,2)="02")+(substr(classe12,1,2)="02")+(substr(classe11,1,2)="02")
+ (substr(classe10,1,2)="02")+(substr(classe09,1,2)="02")+(substr(classe08,1,2)="02")
+ (substr(classe07,1,2)="02")+(substr(classe06,1,2)="02")+(substr(classe05,1,2)="02");

Nb5eme=(substr(classe16,1,2)="03")+(substr(classe15,1,2)="03")+(substr(classe14,1,2)="03")
+ (substr(classe13,1,2)="03")+(substr(classe12,1,2)="03")+(substr(classe11,1,2)="03")
+ (substr(classe10,1,2)="03")+(substr(classe09,1,2)="03")+(substr(classe08,1,2)="03")
+ (substr(classe07,1,2)="03")+(substr(classe06,1,2)="03")+(substr(classe05,1,2)="03");

Nb4eme=(substr(classe16,1,2)="04")+(substr(classe15,1,2)="04")+(substr(classe14,1,2)="04")
+ (substr(classe13,1,2)="04")+(substr(classe12,1,2)="04")+(substr(classe11,1,2)="04")
+ (substr(classe10,1,2)="04")+(substr(classe09,1,2)="04")+(substr(classe08,1,2)="04")
+ (substr(classe07,1,2)="04")+(substr(classe06,1,2)="04")+(substr(classe05,1,2)="04");

NbSegpa=(substr(classe16,1,2)="06")+(substr(classe15,1,2)="06")+(substr(classe14,1,2)="06")
+ (substr(classe13,1,2)="06")+(substr(classe12,1,2)="06")+(substr(classe11,1,2)="06")
+ (substr(classe10,1,2)="06")+(substr(classe09,1,2)="06")+(substr(classe08,1,2)="06")
+ (substr(classe07,1,2)="06")+(substr(classe06,1,2)="06")+(substr(classe05,1,2)="06");

Nb3eme=(substr(classe16,1,2)="05")+(substr(classe15,1,2)="05")+(substr(classe14,1,2)="05")
+ (substr(classe13,1,2)="05")+(substr(classe12,1,2)="05")+(substr(classe11,1,2)="05")
+ (substr(classe10,1,2)="05")+(substr(classe09,1,2)="05")+(substr(classe08,1,2)="05")
+ (substr(classe07,1,2)="05")+(substr(classe06,1,2)="05")+(substr(classe05,1,2)="05");

Nb2de=(substr(classe16,1,2)="08")+(substr(classe15,1,2)="08")+(substr(classe14,1,2)="08")
+ (substr(classe13,1,2)="08")+(substr(classe12,1,2)="08")+(substr(classe11,1,2)="08")
+ (substr(classe10,1,2)="08")+(substr(classe09,1,2)="08")+(substr(classe08,1,2)="08")
+ (substr(classe07,1,2)="08")+(substr(classe06,1,2)="08")+(substr(classe05,1,2)="08");

Nblere=(substr(classe16,1,2)in ("09","10"))+(substr(classe15,1,2)in ("09","10"))+(substr(classe14,1,2)in ("09","10"))
+ (substr(classe13,1,2)in ("09","10"))+(substr(classe12,1,2)in ("09","10"))+(substr(classe11,1,2)in ("09","10"))
+ (substr(classe10,1,2)in ("09","10"))+(substr(classe09,1,2)in ("09","10"))+(substr(classe08,1,2)in ("09","10"))
+ (substr(classe07,1,2)in ("09","10"))+(substr(classe06,1,2)in ("09","10"))+(substr(classe05,1,2)in ("09","10"));

NbTerm=(substr(classe16,1,2)in ("11","12"))+(substr(classe15,1,2)in ("11","12"))+(substr(classe14,1,2)in ("11","12"))
+ (substr(classe13,1,2)in ("11","12"))+(substr(classe12,1,2)in ("11","12"))+(substr(classe11,1,2)in ("11","12"))

```

```

+ (substr(classe10,1,2) in ("11","12")) + (substr(classe09,1,2) in ("11","12")) + (substr(classe08,1,2) in ("11","12"))
+ (substr(classe07,1,2) in ("11","12")) + (substr(classe06,1,2) in ("11","12")) + (substr(classe05,1,2) in ("11","12"));

```

```

NbPrepa=(substr(classe16,1,2) in ("36","37"))
+ (substr(classe15,1,2) in ("36","37"))
+ (substr(classe14,1,2) in ("36","37"))
+ (substr(classe13,1,2) in ("36","37"))
+ (substr(classe12,1,2) in ("36","37"))
+ (substr(classe11,1,2) in ("36","37"))
+ (substr(classe10,1,2) in ("36","37"))
+ (substr(classe09,1,2) in ("36","37"))
+ (substr(classe08,1,2) in ("36","37"))
+ (substr(classe07,1,2) in ("36","37"))
+ (substr(classe06,1,2) in ("36","37"))
+ (substr(classe05,1,2) in ("36","37"));

```

```

Cl_CAP=(substr(classe16,1,2) in ("14","15","16","17","18","19")
or substr(classe15,1,2) in ("14","15","16","17","18","19")
or substr(classe14,1,2) in ("14","15","16","17","18","19")
or substr(classe13,1,2) in ("14","15","16","17","18","19")
or substr(classe12,1,2) in ("14","15","16","17","18","19")
or substr(classe11,1,2) in ("14","15","16","17","18","19")
or substr(classe10,1,2) in ("14","15","16","17","18","19")
or substr(classe09,1,2) in ("14","15","16","17","18","19")
or substr(classe08,1,2) in ("14","15","16","17","18","19")
or substr(classe07,1,2) in ("14","15","16","17","18","19")
or substr(classe06,1,2) in ("14","15","16","17","18","19")
or substr(classe05,1,2) in ("14","15","16","17","18","19"));

```

```

Cl_BEP=(substr(classe16,1,2) in ("20","21","22")
or substr(classe15,1,2) in ("20","21","22")
or substr(classe14,1,2) in ("20","21","22")
or substr(classe13,1,2) in ("20","21","22")
or substr(classe12,1,2) in ("20","21","22")
or substr(classe11,1,2) in ("20","21","22")
or substr(classe10,1,2) in ("20","21","22")
or substr(classe09,1,2) in ("20","21","22")
or substr(classe08,1,2) in ("20","21","22")
or substr(classe07,1,2) in ("20","21","22")
or substr(classe06,1,2) in ("20","21","22")
or substr(classe05,1,2) in ("20","21","22"));

```

```

Cl_Bacpro=(substr(classe16,1,2) in ("23","24","25","26","27","28")
or substr(classe15,1,2) in ("23","24","25","26","27","28")
or substr(classe14,1,2) in ("23","24","25","26","27","28")
or substr(classe13,1,2) in ("23","24","25","26","27","28")
or substr(classe12,1,2) in ("23","24","25","26","27","28")
or substr(classe11,1,2) in ("23","24","25","26","27","28")
or substr(classe10,1,2) in ("23","24","25","26","27","28")
or substr(classe09,1,2) in ("23","24","25","26","27","28")
or substr(classe08,1,2) in ("23","24","25","26","27","28")
or substr(classe07,1,2) in ("23","24","25","26","27","28")
or substr(classe06,1,2) in ("23","24","25","26","27","28")
or substr(classe05,1,2) in ("23","24","25","26","27","28"));

Cl_BTS=(substr(classe16,1,2) in ("38","39","40","41","42","43")
or substr(classe15,1,2) in ("38","39","40","41","42","43")
or substr(classe14,1,2) in ("38","39","40","41","42","43")
or substr(classe13,1,2) in ("38","39","40","41","42","43")
or substr(classe12,1,2) in ("38","39","40","41","42","43")
or substr(classe11,1,2) in ("38","39","40","41","42","43")
or substr(classe10,1,2) in ("38","39","40","41","42","43")
or substr(classe09,1,2) in ("38","39","40","41","42","43")
or substr(classe08,1,2) in ("38","39","40","41","42","43")
or substr(classe07,1,2) in ("38","39","40","41","42","43")
or substr(classe06,1,2) in ("38","39","40","41","42","43")
or substr(classe05,1,2) in ("38","39","40","41","42","43"));

/**** ATTENTION : COMPTAGE DE CONNAISSANCES DES EXAM FAIT SUR LA BASE D'UNE
OPINION DE LA DEFINITION DES MODALITES DE LA VARIABLE FAUTE DE DOCUMENTATION
(on retient les informations de la ligne principale, soit deux examens possibles
pour les fichiers FAERE 2013 et suivants voir documentation *****/

NbBrevet=(Exam1_16="670")+(Exam1_15="670")+(Exam1_14="670")+(Exam1_13="670")+
(Exam1_12="670")+(Exam1_11="670")+(Exam1_10="670")+(Exam1_09="670")+
(Exam1_08="670")+(Exam1_07="670")+(Exam1_06="670")+(Exam1_05="670")+
(Exam2_16="670")+(Exam2_15="670")+(Exam2_14="670")+(Exam2_13="670"); /* pour le 2e diplome connu*/

NbBac =(Exam1_16="470")+(Exam1_15="470")+(Exam1_14="470")+(Exam1_13="470")+ /*Général*/
(Exam1_12="470")+(Exam1_11="470")+(Exam1_10="470")+(Exam1_09="470")+
(Exam1_08="470")+(Exam1_07="470")+(Exam1_06="470")+(Exam1_05="470")

```

```

+ (Exam1_16="430") + (Exam1_15="430") + (Exam1_14="430") + (Exam1_13="430") + /* techno*/
  (Exam1_12="430") + (Exam1_11="430") + (Exam1_10="430") + (Exam1_09="430") +
  (Exam1_08="430") + (Exam1_07="430") + (Exam1_06="430") + (Exam1_05="430")

+ (Exam1_16="400") + (Exam1_15="400") + (Exam1_14="400") + (Exam1_13="400") + /*pro*/
  (Exam1_12="400") + (Exam1_11="400") + (Exam1_10="400") + (Exam1_09="400") +
  (Exam1_08="400") + (Exam1_07="400") + (Exam1_06="400") + (Exam1_05="400")

+ (Exam2_16="470") + (Exam2_15="470") + (Exam2_14="470") + (Exam2_13="470") + /* chacun des 3 bacs */
  (Exam2_16="430") + (Exam2_15="430") + (Exam2_14="430") + (Exam2_13="430") + /* pour le 2e diplome connu*/
  (Exam2_16="400") + (Exam2_15="400") + (Exam2_14="400") + (Exam2_13="400"); /* depuis 2013*/

NbCAPBEP = (Exam1_16="500") + (Exam1_15="500") + (Exam1_14="500") + (Exam1_13="500") +
  (Exam1_12="500") + (Exam1_11="500") + (Exam1_10="500") + (Exam1_09="500") +
  (Exam1_08="500") + (Exam1_07="500") + (Exam1_06="500") + (Exam1_05="500")

+ (Exam1_16="510") + (Exam1_15="510") + (Exam1_14="510") + (Exam1_13="510") +
  (Exam1_12="510") + (Exam1_11="510") + (Exam1_10="510") + (Exam1_09="510") +
  (Exam1_08="510") + (Exam1_07="510") + (Exam1_06="510") + (Exam1_05="510")

+ (Exam2_16="500") + (Exam2_15="500") + (Exam2_14="500") + (Exam2_13="500") + /* pour le 2e diplome connu*/
  (Exam2_16="510") + (Exam2_15="510") + (Exam2_14="510") + (Exam2_13="510");

Redouble_college=(Nb6eme ge 2 or Nb5eme ge 2 or Nb4eme ge 2 or Nb3eme ge 2);
Redouble_Lycee=(Nb2de ge 2 or Nblere ge 2 or Nbterm ge 2);

***** Pour étude de l'obtention du diplôme de sortie, dans Génération et FAERE ;
If diplome1_16="" then diplome1_16="Vide";
If result1_16="" then result1_16="vv";
If nivsor="" then nivsor="vv";

result1_16ag=substr(result1_16,1,1);

Gene_OBT="Apparié:!!an16!!" //Année Term:!!Anter!!" //Obtention:!!obt;

***** Pour étude du fait d'avoir le bac, dans Génération et FAERE ;
Apparie=(Nb_presence ge 1);

Gene_BAC="Apparié:!!Apparie!!" //Bac_Géné:!!BAC;

*repérage des inscriptions au bac dans FAERE;

```

```
Faer_Ins_BacGen=((Exam1_16="470")+(Exam1_15="470")+(Exam1_14="470")+(Exam1_13="470")+
(Exam2_16="470")+(Exam2_15="470")+(Exam2_14="470")+(Exam2_13="470")+
(Exam1_12="470")+(Exam1_11="470")+(Exam1_10="470")+(Exam1_09="470")+
(Exam1_08="470")+(Exam1_07="470")+(Exam1_06="470")+(Exam1_05="470") ge 1);
```

```
Faer_Ins_BacTec=((Exam1_16="430")+(Exam1_15="430")+(Exam1_14="430")+(Exam1_13="430")+
(Exam2_16="430")+(Exam2_15="430")+(Exam2_14="430")+(Exam2_13="430")+
(Exam1_12="430")+(Exam1_11="430")+(Exam1_10="430")+(Exam1_09="430")+
(Exam1_08="430")+(Exam1_07="430")+(Exam1_06="430")+(Exam1_05="430") ge 1);
```

```
Faer_Ins_BacPro=((Exam1_16="400")+(Exam1_15="400")+(Exam1_14="400")+(Exam1_13="400")+
(Exam2_16="400")+(Exam2_15="400")+(Exam2_14="400")+(Exam2_13="400")+
(Exam1_12="400")+(Exam1_11="400")+(Exam1_10="400")+(Exam1_09="400")+
(Exam1_08="400")+(Exam1_07="400")+(Exam1_06="400")+(Exam1_05="400") ge 1);
```

```
Faer_Ins_TOUS=(Faer_Ins_BacGen+Faer_Ins_BacTec+Faer_Ins_BacPro ge 1);
```

*repérage des réussites au bac dans FAERE;

```
Faer_Obt_BacGen=((Exam1_16="470" and substr(result1_16,1,1)="1")
+(Exam2_16="470" and substr(result2_16,1,1)="1")
+(Exam1_15="470" and substr(result1_15,1,1)="1")
+(Exam2_15="470" and substr(result2_15,1,1)="1")
+(Exam1_14="470" and substr(result1_14,1,1)="1")
+(Exam2_14="470" and substr(result2_14,1,1)="1")
+(Exam1_13="470" and substr(result1_13,1,1)="1")
+(Exam2_13="470" and substr(result2_13,1,1)="1")
+(Exam1_12="470" and substr(result1_12,1,1)="1")
+(Exam1_11="470" and substr(result1_11,1,1)="1")
+(Exam1_10="470" and substr(result1_10,1,1)="1")
+(Exam1_09="470" and substr(result1_09,1,1)="1")
+(Exam1_08="470" and substr(result1_08,1,1)="1")
+(Exam1_07="470" and substr(result1_07,1,1)="1")
+(Exam1_06="470" and substr(result1_06,1,1)="1")
+(Exam1_05="470" and substr(result1_05,1,1)="1") ge 1);
```

```
Faer_Obt_BacTec=((Exam1_16="430" and substr(result1_16,1,1)="1")
+(Exam2_16="430" and substr(result2_16,1,1)="1")
+(Exam1_15="430" and substr(result1_15,1,1)="1")
```



```

+ (Exam2_15="430" and substr(result2_15,1,1)="1")
+ (Exam1_14="430" and substr(result1_14,1,1)="1")
+ (Exam2_14="430" and substr(result2_14,1,1)="1")
+ (Exam1_13="430" and substr(result1_13,1,1)="1")
+ (Exam2_13="430" and substr(result2_13,1,1)="1")
+ (Exam1_12="430" and substr(result1_12,1,1)="1")
+ (Exam1_11="430" and substr(result1_11,1,1)="1")
+ (Exam1_10="430" and substr(result1_10,1,1)="1")
+ (Exam1_09="430" and substr(result1_09,1,1)="1")
+ (Exam1_08="430" and substr(result1_08,1,1)="1")
+ (Exam1_07="430" and substr(result1_07,1,1)="1")
+ (Exam1_06="430" and substr(result1_06,1,1)="1")
+ (Exam1_05="430" and substr(result1_05,1,1)="1") ge 1);

```

```

Faer_Obt_BacPro=( (Exam1_16="400" and substr(result1_16,1,1)="1")
+ (Exam2_16="400" and substr(result2_16,1,1)="1")
+ (Exam1_15="400" and substr(result1_15,1,1)="1")
+ (Exam2_15="400" and substr(result2_15,1,1)="1")
+ (Exam1_14="400" and substr(result1_14,1,1)="1")
+ (Exam2_14="400" and substr(result2_14,1,1)="1")
+ (Exam1_13="400" and substr(result1_13,1,1)="1")
+ (Exam2_13="400" and substr(result2_13,1,1)="1")
+ (Exam1_12="400" and substr(result1_12,1,1)="1")
+ (Exam1_11="400" and substr(result1_11,1,1)="1")
+ (Exam1_10="400" and substr(result1_10,1,1)="1")
+ (Exam1_09="400" and substr(result1_09,1,1)="1")
+ (Exam1_08="400" and substr(result1_08,1,1)="1")
+ (Exam1_07="400" and substr(result1_07,1,1)="1")
+ (Exam1_06="400" and substr(result1_06,1,1)="1")
+ (Exam1_05="400" and substr(result1_05,1,1)="1") ge 1);

```

```

Faer_Obt_TOUS=(Faer_obt_BacGen+Faer_Obt_BacTec+Faer_Obt_BacPro ge 1);
Nb_Faer_Obt=Faer_obt_BacGen+Faer_Obt_BacTec+Faer_Obt_BacPro;

```

```

* repérage de l'année du bac dans FAERE;

```

```

Faer_Anbac=max(pres16*(substr(result1_16,1,1)="1")*(Exam1_16 in ("470","430","400")),
pres16*(substr(result2_16,1,1)="1")*(Exam2_16 in ("470","430","400")),
pres15*(substr(result1_15,1,1)="1")*(Exam1_15 in ("470","430","400")),
pres15*(substr(result2_15,1,1)="1")*(Exam2_15 in ("470","430","400")),
pres14*(substr(result1_14,1,1)="1")*(Exam1_14 in ("470","430","400")),
pres14*(substr(result2_14,1,1)="1")*(Exam2_14 in ("470","430","400")),
pres13*(substr(result1_13,1,1)="1")*(Exam1_13 in ("470","430","400")),

```

```

pres13*(substr(result2_13,1,1)="1")*(Exam2_13 in ("470","430","400")),
pres12*(substr(result1_12,1,1)="1")*(Exam1_12 in ("470","430","400")),
pres11*(substr(result1_11,1,1)="1")*(Exam1_11 in ("470","430","400")),
pres10*(substr(result1_10,1,1)="1")*(Exam1_10 in ("470","430","400")),
pres09*(substr(result1_09,1,1)="1")*(Exam1_09 in ("470","430","400")),
pres08*(substr(result1_08,1,1)="1")*(Exam1_08 in ("470","430","400")),
pres07*(substr(result1_07,1,1)="1")*(Exam1_07 in ("470","430","400")),
pres06*(substr(result1_06,1,1)="1")*(Exam1_06 in ("470","430","400")),
pres05*(substr(result1_05,1,1)="1")*(Exam1_05 in ("470","430","400")));

If Faer_Anbac=Anbac*1-1 then Ecart_Anbac="Même année";
Else If Faer_Anbac=Anbac*1 then Ecart_Anbac="Ecart 1 an";
Else If Faer_Anbac=Anbac*1-2 then Ecart_Anbac="Ecart 1 an";
Else Ecart_Anbac="Plus ";

If bac_type='1' and Faer_Obt_BacGen=1 then Ecart_Typbac="Même type";
Else If bac_type='2' and Faer_Obt_BacTec=1 then Ecart_Typbac="Même type";
Else If bac_type='3' and Faer_Obt_BacPro=1 then Ecart_Typbac="Même type";
Else Ecart_Typbac="Différent";

BacCoherent=(Ecart_Anbac="Même année" and Ecart_Typbac="Même type");

if bac_mention="" then bac_mention="v";

*repérage de la mention du bac dans FAERE;
Faer_mentionbac="vv";
if Faer_Anbac=2016 and Exam1_16 in ("470","430","400") then Faer_mentionbac=result1_16;
else if Faer_Anbac=2016 and Exam2_16 in ("470","430","400") then Faer_mentionbac=result2_16;
if Faer_Anbac=2015 and Exam1_15 in ("470","430","400") then Faer_mentionbac=result1_15;
else if Faer_Anbac=2015 and Exam2_15 in ("470","430","400") then Faer_mentionbac=result2_15;
if Faer_Anbac=2014 and Exam1_14 in ("470","430","400") then Faer_mentionbac=result1_14;
else if Faer_Anbac=2014 and Exam2_14 in ("470","430","400") then Faer_mentionbac=result2_14;
if Faer_Anbac=2013 and Exam1_13 in ("470","430","400") then Faer_mentionbac=result1_13;
else if Faer_Anbac=2013 and Exam2_13 in ("470","430","400") then Faer_mentionbac=result2_13;
if Faer_Anbac=2012 and Exam1_12 in ("470","430","400") then Faer_mentionbac=result1_12;
if Faer_Anbac=2011 and Exam1_11 in ("470","430","400") then Faer_mentionbac=result1_11;
if Faer_Anbac=2010 and Exam1_10 in ("470","430","400") then Faer_mentionbac=result1_10;
if Faer_Anbac=2009 and Exam1_09 in ("470","430","400") then Faer_mentionbac=result1_09;
if Faer_Anbac=2008 and Exam1_08 in ("470","430","400") then Faer_mentionbac=result1_08;
if Faer_Anbac=2007 and Exam1_07 in ("470","430","400") then Faer_mentionbac=result1_07;
if Faer_Anbac=2006 and Exam1_06 in ("470","430","400") then Faer_mentionbac=result1_06;
if Faer_Anbac=2005 and Exam1_05 in ("470","430","400") then Faer_mentionbac=result1_05;

```

```

*repérage de la source d'info (SySCA/SIFA) du bac et du lieu de résidence
au bac dans FAERE;
if Faer_Anbac=2016 then do; Faer_SourcBac=Source_16;
    dptbac=dpt16;acaBac=acadresid16;end;
if Faer_Anbac=2015 then do; Faer_SourcBac=Source_15;
    dptbac=dpt15;acaBac=acadresid15;end;
if Faer_Anbac=2014 then do; Faer_SourcBac=Source_14;
    dptbac=dpt14;acaBac=acadresid14;end;
if Faer_Anbac=2013 then do; Faer_SourcBac=Source_13;
    dptbac=dpt13;acaBac=acadresid13;end;
if Faer_Anbac=2012 then do; Faer_SourcBac=Source_12;
    dptbac=dpt12;acaBac=acadresid12;end;
if Faer_Anbac=2011 then do; Faer_SourcBac=Source_11;
    dptbac=dpt11;acaBac=acadresid11;end;
if Faer_Anbac=2010 then do; Faer_SourcBac=Source_10;
    dptbac=dpt10;acaBac=acadresid10;end;
if Faer_Anbac=2009 then do; Faer_SourcBac=Source_09;
    dptbac=dpt09;acaBac=acadresid09;end;
if Faer_Anbac=2008 then do; Faer_SourcBac=Source_08;
    dptbac=dpt08;acaBac=acadresid08;end;
if Faer_Anbac=2007 then do; Faer_SourcBac=Source_07;
    dptbac=dpt07;acaBac=acadresid07;end;
if Faer_Anbac=2006 then do; Faer_SourcBac=Source_06;
    dptbac=dpt06;acaBac=acadresid06;end;
if Faer_Anbac=2005 then do; Faer_SourcBac=Source_05;
    dptbac=dpt05;acaBac=acadresid05;end;

***** Pour étude redoublement au collège;
If PS050 = "" then PS050="9";* question non posée;

RedoubCollCoher="App: "!!Apparie!!"// FAERE: "!!Redouble_college!!"// Génération: "!!PS050;

***** Pour étude d'un passage en 3e;
If PS040 = "" then PS040="9";* question non posée;
If PS040= "4" then PS040="0";* pas de 3ème;

If nb3eme=0 then Faer_3e_gen =0;
Else if nb3eme=1 then Faer_3e_gen =1;
Else if nb3eme ge 2 then Faer_3e_gen =2;
Else Faer_3e_gen=9;

```

```
Cohér3eme="App:!!Apparie!!"// FAERE:!!Faer_3e_gen!!"//Génération: "!!PS040;
```

```
***** Classe suivie après la 3ème et lieu résidence en 3e;
```

```
Derniere3e=Max(pres16*(substr(classe16,1,2)="05"),
               pres15*(substr(classe15,1,2)="05"),
               pres14*(substr(classe14,1,2)="05"),
               pres13*(substr(classe13,1,2)="05"),
               pres12*(substr(classe12,1,2)="05"),
               pres11*(substr(classe11,1,2)="05"),
               pres10*(substr(classe10,1,2)="05"),
               pres09*(substr(classe09,1,2)="05"),
               pres08*(substr(classe08,1,2)="05"),
               pres07*(substr(classe07,1,2)="05"),
               pres06*(substr(classe06,1,2)="05"),
               pres05*(substr(classe05,1,2)="05"));
```

```
ClassePost3e="99-Pas d'info FAERE ";
```

```
If      Derniere3e=2016 then do;
    ClassePost3e="99-3e en 2016 ";
    dpt3e=dpt16; Aca3e=AcadResid16;end;
else If Derniere3e=2015 then do;
    ClassePost3e=Classe16;dpt3e=dpt15; Aca3e=AcadResid15;end;
else If Derniere3e=2014 then do;
    ClassePost3e=Classe15;dpt3e=dpt14; Aca3e=AcadResid14;end;
else If Derniere3e=2013 then do;
    ClassePost3e=Classe14;dpt3e=dpt13; Aca3e=AcadResid13;end;
else If Derniere3e=2012 then do;
    ClassePost3e=Classe13;dpt3e=dpt12; Aca3e=AcadResid12;end;
else If Derniere3e=2011 then do;
    ClassePost3e=Classe12;dpt3e=dpt11; Aca3e=AcadResid11;end;
else If Derniere3e=2010 then do;
    ClassePost3e=Classe11;dpt3e=dpt10; Aca3e=AcadResid10;end;
else If Derniere3e=2009 then do;
    ClassePost3e=Classe10;dpt3e=dpt09; Aca3e=AcadResid09;end;
else If Derniere3e=2008 then do;
    ClassePost3e=Classe09;dpt3e=dpt08; Aca3e=AcadResid08;end;
else If Derniere3e=2007 then do;
    ClassePost3e=Classe08;dpt3e=dpt07; Aca3e=AcadResid07;end;
else If Derniere3e=2006 then do ;
    ClassePost3e=Classe07;dpt3e=dpt06; Aca3e=AcadResid06;end;
else If Derniere3e=2005 then do;
    ClassePost3e=Classe06;dpt3e=dpt05; Aca3e=AcadResid05;end;
```

```

If ClassePost3e=" " then ClassePost3e="98-Pas d'info FAERE " ; ;
If PS060="" then PS060="V";

***** Lieu de résidence en 6ème;
Derniere6e=Max(pres16*(substr(classe16,1,2)="02"),
               pres15*(substr(classe15,1,2)="02"),
               pres14*(substr(classe14,1,2)="02"),
               pres13*(substr(classe13,1,2)="02"),
               pres12*(substr(classe12,1,2)="02"),
               pres11*(substr(classe11,1,2)="02"),
               pres10*(substr(classe10,1,2)="02"),
               pres09*(substr(classe09,1,2)="02"),
               pres08*(substr(classe08,1,2)="02"),
               pres07*(substr(classe07,1,2)="02"),
               pres06*(substr(classe06,1,2)="02"),
               pres05*(substr(classe05,1,2)="02"));

If      Derniere6e=2016 then do; dpt6e=dpt16; Aca6e=AcadResid16;end;
else If Derniere6e=2015 then do; dpt6e=dpt15; Aca6e=AcadResid15;end;
else If Derniere6e=2014 then do; dpt6e=dpt14; Aca6e=AcadResid14;end;
else If Derniere6e=2013 then do; dpt6e=dpt13; Aca6e=AcadResid13;end;
else If Derniere6e=2012 then do; dpt6e=dpt12; Aca6e=AcadResid12;end;
else If Derniere6e=2011 then do; dpt6e=dpt11; Aca6e=AcadResid11;end;
else If Derniere6e=2010 then do; dpt6e=dpt10; Aca6e=AcadResid10;end;
else If Derniere6e=2009 then do; dpt6e=dpt09; Aca6e=AcadResid09;end;
else If Derniere6e=2008 then do; dpt6e=dpt08; Aca6e=AcadResid08;end;
else If Derniere6e=2007 then do; dpt6e=dpt07; Aca6e=AcadResid07;end;
else If Derniere6e=2006 then do; dpt6e=dpt06; Aca6e=AcadResid06;end;
else If Derniere6e=2005 then do; dpt6e=dpt05; Aca6e=AcadResid05;end;

*** Comparaison des départements de résidence en 6e;
Dpt6eCereq=substr(SIXIEMEDEP,1,2); * il y a du "***" et "NC" ;
If Dpt6eCereq in ("**","NC") then Dpt6eCereq="ZZ";

If Dpt6e="$ " then Dpt6e="ZZ";
If DptBac="$ " then DptBac="ZZ";

CompaDpt6e="Autres cas ";
If dpt6e=Dpt6eCereq and dpt6e ne "" and dpt6e ne "ZZ"
then CompaDpt6e="Même département";

```

```

else If dpt6e=Dpt6eCereq and dpt6e= "ZZ"
  then CompaDpt6e="Les 2 à ZZ ";
else If dpt6e ne Dpt6eCereq and dpt6e = "ZZ" and Dpt6eCereq ne ""
  then CompaDpt6e="Un des 2 à ZZ ";
else If dpt6e ne Dpt6eCereq and dpt6e ne "" and Dpt6eCereq="ZZ"
  then CompaDpt6e="Un des 2 à ZZ ";
else If dpt6e ne "" and Dpt6eCereq ne "" and dpt6e ne Dpt6eCereq
  then CompaDpt6e="Les 2 différents";

*** Comparaison des départements de résidence au Bac;
dptBacCereq=substr(BACDEP,1,2);
If dptBacCereq in ("**","NC") then dptBacCereq="ZZ";

CompaDptBac="Autres cas ";
If dptBac=DptBacCereq and dptBac ne "" and dptBac ne "ZZ"
  then CompaDptBac="Même département";
else If dptBac=DptBacCereq and dptBac= "ZZ"
  then CompaDptBac="Les 2 à ZZ ";
else If dptBac ne DptBacCereq and dptBac = "ZZ" and DptBacCereq ne ""
  then CompaDptBac="Un des 2 à ZZ ";
else If dptBac ne DptBacCereq and dptBac ne "" and DptBacCereq="ZZ"
  then CompaDptBac="Un des 2 à ZZ ";
else If dptBac ne "" and DptBacCereq ne "" and dptBac ne DptBacCereq
  then CompaDptBac="Les 2 différents";

*** Changement d'académies entre la 6e et la 3e;
If Aca6e ne "ZZZZZZZZ" and Aca6e ne "" and Aca3e ne "ZZZZZZZZ" and Aca3e ne ""
  then stabAcad6e3e=(Aca6e=Aca3e);
else stabAcad6e3e=9;

*** Changement d'académies entre la 6e et le bac;
If Aca6e ne "ZZZZZZZZ" and Aca6e ne "" and Acabac ne "ZZZZZZZZ" and Acabac ne ""
  then stabAcad6eBac=(Aca6e=Acabac);
else stabAcad6ebac=9;

**** Selection pour étude de non réponse;
SelecNonRep =(substr(Classe16,1,2) in ("02","03","04","05","06","07","08","09","10",
    "11","12","13","14","15","16","17","18","19","20","21","22","23","24"));

**** pays de naissance et nationalité dans Génération;
If SD040 in ("1","2") then SD040=SD040;

```

```

Else if Lieunais="00" then SD040="0";
Else SD040="Z";
If Natio16="" then Natio16="VVVVVV";
run;

Proc freq data=historique;

tables (PrefixIdent phd_qeef2 Nivsor)*Nb_presence/nopercent nocol norow;
run;

Title "3.Nombre de présences dans les fichiers annuels de la DEPP des répondants";
Proc freq data=historique;
tables Nb_presence*Rep2017Diff/nopercent norow nocol;
tables (PrefixIdent phd_qeef2 Nivsor)*Nb_presence/nopercent nocol norow;
run;

Title "4.Nombre de présence dans les classes de college et lycée généraux/techno";
Title2 "(Champ des IDENT au moins une fois appariée avec DEPP)";
Proc freq data=historique (where=(Nb_presence ge 1));
tables (PrefixIdent phd_qeef2 Nivsor)*(NbPrim Nb6eme Nb5eme Nb4eme Nb3eme Nb2de Nb1ere Nbterm NbPrepa)/nopercent nocol
norow;
run;

Proc freq data=historique (where=(Nb_presence ge 1));
tables (PrefixIdent phd_qeef2 Nivsor)*(Redouble_college Redouble_Lycee NbSegpa)/nopercent nocol norow;
run;

Title "4. Nombre d'individus passés au moins une année dans la classe indiquée";
Title2 "(Champ des IDENT au moins une fois appariée avec DEPP)";
Proc freq data=historique (where=(Nb_presence ge 1)); ;
tables (PrefixIdent phd_qeef2 Nivsor)*(Cl_CAP Cl_BEP Cl_Bacpro Cl_BTS)/nopercent nocol norow;
run;

Title "4bis.INFO NON PRINCIPALE Nombre d'individus par classe une année données (lignes 2 et 3)";
Title2 "(Champ des IDENT au moins une fois appariée avec DEPP)";
Proc freq data=historique (where=(Nb_presence ge 1)); ;
tables      classe16_L2 classe16_L3 classe15_L2 classe15_L3
             classe14_L2 classe14_L3 classe13_L2 classe13_L3
             classe12_L2 classe12_L3 classe11_L2 classe11_L3
             classe10_L2 classe10_L3 classe09_L2 classe09_L3
             classe08_L2 classe08_L3 classe07_L2 classe07_L3
             classe06_L2 classe06_L3 classe05_L2 classe05_L3;

```

```
run;
```

```
Title "5.Nombre d'inscription à examen (ligne principale ONLY ; 2 exam. possibles depuis 2013)";
```

```
Title2 "(Champ des IDENT au moins une fois appariée avec DEPP)";
```

```
Proc freq data=historique (where=(Nb_presence ge 1)); ;
```

```
tables (PrefixIdent phd_qeef2 Nivsor)*(NbBrevet NbBAC NbCAPBEP)/nopercent nocol norow;
```

```
run;
```

```
Title2 "Quels examens sont faits en parallèle (étudiable depuis 2013)?";
```

```
Proc freq data=historique (where=(Nb_presence ge 1)); ;
```

```
tables Exam1_16*Exam2_16 Exam1_15*Exam2_15 Exam1_14*Exam2_14 Exam1_13*Exam2_13 /nopercent nocol norow;
```

```
run;
```

```
Title "6.Nombre d'années repérés dans le secteur privé";
```

```
Title2 "(Champ des IDENT au moins une fois appariée avec DEPP)";
```

```
Proc freq data=historique (where=(Nb_presence ge 1));
```

```
tables (PrefixIdent phd_qeef2 Nivsor)*(NbPrive)/nopercent nocol norow;
```

```
tables (PrefixIdent phd_qeef2 Nivsor)*(NbPublic)/nopercent nocol norow;
```

```
run;
```

```
Title "6.Nombre d'années repérés dans le secteur privé-PONDERE";
```

```
Title2 "(Champ des IDENT au moins une fois appariée avec DEPP)";
```

```
Proc freq data=historique (where=(Nb_presence ge 1));
```

```
tables phd_qeef2 *(NbPrive NbPublic)/nopercent nocol norow;
```

```
weight ponddef;
```

```
run;
```

```
Title "7. Première et dernière année de presence dans FAERE";
```

```
Title2 "(Champ des IDENT au moins une fois appariée avec DEPP)";
```

```
Proc freq data=historique (where=(Nb_presence ge 1));
```

```
tables (PrefixIdent phd_qeef2 Nivsor)*(Firstyear Lastyear)/nopercent nocol norow;
```

```
run;
```

```
Title "7. Nombre d'absences entre la première et dernière année de presence dans FAERE";
```

```
Title2 "(Champ des IDENT au moins une fois appariée avec DEPP)";
```

```
Proc freq data=historique (where=(Nb_presence ge 1));
```

```
tables (PrefixIdent phd_qeef2 Nivsor)*Nbtrous/nopercent nocol norow;
```

```
run;
```

```
Title "8. Première et dernière classe connue dans FAERE";
```

```
Title2 "(Champ des IDENT au moins une fois appariée avec DEPP)";
```

```
Proc freq data=historique (where=(Nb_presence ge 1));
```



```

tables (PrefixIdent phd_qeef2 Nivsor)*(Firstclasse Lastclasse)/nopercent nocol norow;
run;

*****;
* Comparaisons de l'obtention du diplôme de sortie entre Génération et FAERE2016;
*****;

Title "9. Comparaison Génération/FAERE pour la variable OBT de Génération (obtention du diplôme de sortie)";
Title2 "champ : tous";
Proc freq data=historique;
tables Gene_OBT*rep2017diff /nopercent nocol norow;
tables nivsor*(obt diplome1_16 prefixIDENT) /nopercent nocol norow;
run;

* CAP/BEP;
Title2 "Niveau de sortie CAP/BEP dans Génération";
Proc freq data=historique (where=(nivsor='02'));
tables Gene_OBT*(obt diplome1_16) /nopercent nocol norow;
run;
Title2 "Niveau de sortie CAP/BEP dans Génération et Diplome1_16=CEP/BEP";
Proc freq data=historique (where=(nivsor='02' and Diplome1_16="CAP/BEP "));
tables Gene_OBT*(result1_16ag) /nopercent nocol norow;
run;

Title2 "PONDERE-Niveau de sortie CAP/BEP dans Génération";
Proc freq data=historique (where=(nivsor='02'));
tables Gene_OBT*(obt diplome1_16) /nopercent nocol norow;weight pondef;
run;
Title2 "PONDERE-Niveau de sortie CAP/BEP dans Génération et Diplome1_16=CEP/BEP";
Proc freq data=historique (where=(nivsor='02' and Diplome1_16="CAP/BEP "));
tables Gene_OBT*(result1_16ag) /nopercent nocol norow;weight pondef;
run;

* Bac Pro;
Title2 "Niveau de sortie Bac Pro dans Génération";
Proc freq data=historique (where=(nivsor='03'));
tables Gene_OBT*(obt diplome1_16) /nopercent nocol norow;
run;
Title2 "Niveau de sortie Bac Pro dans Génération et Diplome1_16=Bac pro.";
Proc freq data=historique (where=(nivsor='03' and Diplome1_16="Bac pro. "));
tables Gene_OBT*(result1_16ag) /nopercent nocol norow;
run;

```

```

Title2 "PONDERE-Niveau de sortie Bac Pro dans Génération";
Proc freq data=historique (where=(nivsor='03'));
tables Gene_OBT*(obt diplome1_16) /nopercent nocol norow;weight pondef;
run;

Title2 "PONDERE-Niveau de sortie Bac Pro dans Génération et Diplome1_16=Bac pro.";
Proc freq data=historique (where=(nivsor='03' and Diplome1_16="Bac pro. "));
tables Gene_OBT*(result1_16ag) /nopercent nocol norow;weight pondef;
run;

* Bac techno;
Title2 "Niveau de sortie Bac techno dans Génération";
Proc freq data=historique (where=(nivsor='05'));
tables Gene_OBT*(obt diplome1_16) /nopercent nocol norow;
run;
Title2 "Niveau de sortie Bac techno dans Génération et Diplome1_16=Bac techno";
Proc freq data=historique (where=(nivsor='05' and Diplome1_16="Bac techno"));
tables Gene_OBT*(result1_16ag) /nopercent nocol norow;
run;

* Bac général;
Title2 "Niveau de sortie Bac général dans Génération";
Proc freq data=historique (where=(nivsor='06'));
tables Gene_OBT*(obt diplome1_16) /nopercent nocol norow;
run;
Title2 "Niveau de sortie Bac général dans Génération et Diplome1_16=Bac Géné";
Proc freq data=historique (where=(nivsor='06' and Diplome1_16="Bac Géné. "));
tables Gene_OBT*(result1_16ag) /nopercent nocol norow;
run;

*****;
* Comparaisons de l'obtention d'un bac entre Génération et l'ensemble des fichiers FAERE;
*****;

Title "10. Comparaison Génération/FAERE pour l'obtention d'un bac";
Title2 "Avoir le Bac dans Génération (BAC) et avoir été inscrit dans FAERE (FAER_INS...)";
Proc freq data=historique;
tables Gene_BAC*rep2017diff /nopercent nocol norow;
tables Nivsor*Bac /nopercent nocol norow;
tables Gene_BAC*(Faer_Ins_BacGen Faer_Ins_BacTec Faer_Ins_BacPro Faer_Ins_TOUS) /nopercent nocol norow;
tables Gene_BAC*(Faer_Obt_BacGen Faer_Obt_BacTec Faer_Obt_BacPro Faer_Obt_TOUS) /nopercent nocol norow;
run;

```

```
Title2 "PONDERE-Avoir le Bac dans Génération (BAC) et avoir été inscrit dans FAERE (FAER_INS...)";
Proc freq data=historique;
tables Gene_BAC*(Faer_Ins_BacGen Faer_Ins_BacTec Faer_Ins_BacPro Faer_Ins_TOUS) /nopercent nocol norow;
tables Gene_BAC*(Faer_Obt_BacGen Faer_Obt_BacTec Faer_Obt_BacPro Faer_Obt_TOUS) /nopercent nocol norow;
weight ponddef;
run;

Title2 "CHAMP RESTREINT AUX AYANT LE BAC DANS GENERATION";
Proc freq data=historique (where=(bac="1"));
tables (Nivsor PrefixIdent)*Faer_Obt_TOUS /nopercent nocol norow;
run;

Title2 "PONDERE-CHAMP RESTREINT AUX AYANT LE BAC DANS GENERATION";
Proc freq data=historique (where=(bac="1"));
tables (Nivsor PrefixIdent)*Faer_Obt_TOUS /nopercent nocol norow;
weight ponddef;
run;

Title2 "Cohérences des type de bac (CHAMP RESTREINT AUX AYANT LE BAC DANS GENERATION & 1 seul dans FAERE)";
Proc freq data=historique (where=(bac="1" and Nb_Faer_Obt =1));
tables Bac_type*(Faer_Obt_BacGen Faer_Obt_BacTec Faer_Obt_BacPro)/nopercent nocol norow;

run;
*** examen des incohérences pour le bac pro;
Data verif; set historique (where=(bac="1" and Nb_Faer_Obt =1));
If Faer_Obt_BacPro=1 and Bac_Type="3" then OK=1;
else If Faer_Obt_BacPro=1 and Bac_Type ne"3" then OK=0;
else delete;
run;
Proc freq data=verif;
tables Bac_type*Faer_Obt_BacPro/nopercent norow;
tables mode*Bac_type*Faer_Obt_BacPro/nopercent norow;
tables OK;
tables nivsor*OK/nopercent nocol nofreq;
tables nivsor*OK/nopercent nocol norow;
run;
Data verif; set _null_;
run;
****;

Title2 "Cohérences des type de bac (CHAMP RESTREINT AUX AYANT LE BAC DANS GENERATION & 2 ou + dans FAERE)";
```

```

Proc freq data=historique (where=(bac="1" and Nb_Faer_Obt ge 2));
tables Bac_type*(Faer_Obt_BacGen Faer_Obt_BacTec Faer_Obt_BacPro)/nopercent nocol norow;
run;

Title2 "Cohérences des années du de bac (CHAMP RESTREINT AUX AYANT LE BAC DANS GENERATION & 1 seul dans FAERE)";
Proc freq data=historique (where=(bac="1" and Nb_Faer_Obt =1));
tables anbac*Faer_Anbac/nopercent nocol norow;
tables Ecart_Anbac*Ecart_Typbac/nopercent nocol norow;
Tables BacCoherent (Bac_Type PrefixIdent phd_qeef2 nivsor)*bacCoherent/nopercent nocol norow;
run;

Title2 "PONDERE (CHAMP RESTREINT AUX AYANT LE BAC DANS GENERATION & 1 seul dans FAERE)";
Proc freq data=historique (where=(bac="1" and Nb_Faer_Obt =1));
Tables BacCoherent (Bac_Type PrefixIdent phd_qeef2)*bacCoherent/nopercent nocol norow;
Tables BacCoherent*Faer_SourcBac /nopercent norow nofreq;
weight pondef;
run;

Title2 "NON PONDERE- Les dpts du BAC dans les 2 sources";
Proc freq data=historique;
tables CompadptBac;
tables CompaDptBac*(bac Rep2017Diff)/nopercent nocol norow;
run;

*****;
* Mention bac;
*****;

Title "10b. Comparaison Génération/FAERE pour la mention du bac";
Title2 "(CHAMP RESTREINT AUX AYANT LE BAC DANS GENERATION & 1 seul dans FAERE)";
Proc freq data=historique (where=(bac="1" and Nb_Faer_Obt =1));
tables bac_mention*Faer_mentionbac/nopercent nocol norow;
tables mode*bac_mention*Faer_mentionbac/nopercent nocol norow;
run;

Title2 "PONDERE-(CHAMP RESTREINT AUX AYANT LE BAC DANS GENERATION & 1 seul dans FAERE)";
Proc freq data=historique (where=(bac="1" and Nb_Faer_Obt =1));
tables bac_mention*Faer_mentionbac/nopercent nocol norow;
weight pondef;
run;
Title2 "PONDERE-(CHAMP RESTREINT AUX AYANT LE BAC DANS GENERATION & 1 seul dans FAERE & BACCOHERENT=1)";

```

```
Proc freq data=historique (where=(bac="1" and Nb_Faer_Obt =1 and BacCoherent=1));
tables bac_mention*Faer_mentionbac/nopercent nocol norow;
weight ponddef;
run;
```

```
*****;
* Redoublement collège;
*****;
Title "11. Comparaison Génération/FAERE pour le redoublement au collège";
Title2 "PS050 dans Génération (champ : niveau d'études de 3e à Bac+3)";
Proc freq data=historique;
tables (PrefixIDENT NivSor PHD_QEEF2)*PS050 /nopercent nocol norow;
tables Redouble_college*PS050 /nopercent nocol norow;
run;
```

```
Title2 "PONDERE-PS050 dans Génération (champ : niveau d'études de 3e à Bac+3)";
Proc freq data=historique;
tables Redouble_college*PS050 /nopercent nocol norow;
tables (PrefixIDENT NivSor PHD_QEEF2)*RedoubCollCoher/nopercent nocol norow;
weight ponddef;
run;
```

```
*** comparaison des département de 6e;
Title2 "NON PONDERE- Les dpts de 6e dans les 2 sources";
Proc freq data=historique (where=(ps020 in ("1","2")));
tables CompaDpt6e*Rep2017Diff/nopercent nocol norow;
tables CompaDpt6e*nb6eme/nopercent nocol norow;
run;
```

```
*****;
* Avoir suivi une 3e;
*****;
Title "12. Comparaison Génération/FAERE pour le fait d'avoir suivi une 3e";
Title2 "PS040 dans Génération (champ : niveau d'études de 3e à Bac+3)";
Proc freq data=historique;
tables (PrefixIDENT NivSor PHD_QEEF2)*Coher3eme /nopercent nocol norow;
run;
```

```
Title2 "PONDERE- (champ : niveau d'études de 3e à Bac+3)";
Proc freq data=historique;
```

```

tables (PrefixIDENT NivSor PHD_QEEF2)*Coher3eme /nopercent nocol norow;
tables Nb3eme*PS040 /nopercent nocol norow;
weight PONDEF;
run;

*****;
* Classe post3e;
*****;

Title "13. Comparaison Génération/FAERE sur la classe suivie après la 3e";
Title2 "PS060 dans Génération (champ : niveau d'études de 3e à Bac+3 & avoir suivi 3e)";
Proc freq data=historique;
tables ClassePost3e;
tables ClassePost3e*nbs3eme/nopercent nocol norow;
tables ClassePost3e*PS060/nopercent nocol norow;
run;

Title "13. Comparaison Génération/FAERE sur la classe suivie après la 3e";
Title2 "PONDERE (champ : niveau d'études de 3e à Bac+3 & avoir suivi 3e)";
Proc freq data=historique;
tables ClassePost3e*nbs3eme/nopercent nocol norow;
tables ClassePost3e*PS060/nopercent nocol norow;
weight pondef;
run;

Proc freq data=historique (where=(substr(nivsor,1,2) in ("02","03")));
tables nivsor*PS060/nopercent nocol norow;
tables mode*nivsor*PS060/nopercent nocol norow;
weight pondef;
run;

*****;
*****;
*****;
* Non répondants et repondants;
*****;
*****;
*****;

Title "14. Non réponse des SYSCA & SIFA";

```

```

Title2 "Comparaison NIVSOR et SELECNONREP";
Proc freq data=historique (where=(rep2017diff=1));
tables nivsor*SelecNonRep / nopercnt nocol;
tables nivsor*SelecNonRep / nopercnt norow;
tables prefixIDENT*SelecNonRep / nopercnt norow nocol;
run;

Title2 "Champ des Classe16 correspondant supposé à NIVSOR = 1,2,3,5,6";
Proc freq data=historique (where=(SelecNonRep=1));
tables Classe16*rep2017diff/nopercnt norow nocol;
tables Redouble_college*rep2017diff/nopercnt norow nocol;
tables Redouble_Lycee*rep2017diff/nopercnt norow nocol;
tables nbbrevet*rep2017diff/nopercnt norow nocol;
tables boursier16*rep2017diff/nopercnt norow nocol;
tables CS16*rep2017diff/nopercnt norow nocol;
tables natio16*rep2017diff/nopercnt norow nocol;
run;

Title2 "LOGISTIQUE-Champ des retrouvés dans FAERE2016";
Proc logistic data=historique (where=(SelecNonRep=1));
Class classe16 Redouble_college Redouble_Lycee nbbrevet boursier16 cs16 natio16;
model rep2017diff=classe16 Redouble_college Redouble_Lycee nbbrevet boursier16 cs16 natio16
/selection=stepwise;
run;

*****;
*****;
* 15.COMPLEMENTS
*****;
*****;

Title "15a. Analyse de la nationalité dans Génération (SD040) et FAERE (NATIO&AN)";
Title2 "Non pondéré";
title3 "SYSCA ET SIFA (pas de SIFA avant 2008) ";
Proc freq data=historique;
tables lieunais*SD040/nopercnt norow nocol;

```

```
tables natio16*sd040/nopercent norow nocol;
run;

Proc freq data=historique;
tables natio16*natio15/nopercent norow nocol;
tables natio15*natio14/nopercent norow nocol;
tables natio14*natio13/nopercent norow nocol;
tables natio13*natio12/nopercent norow nocol;
tables natio12*natio11/nopercent norow nocol;
tables natio11*natio10/nopercent norow nocol;
tables natio16*(natio14 natio13 natio12 natio11 natio10)/nopercent norow nocol;
tables natio16*(natio09 natio08 natio07 natio06 natio05)/nopercent norow nocol;
run;

title3 "SYSCA uniquement les deux années";
Proc freq data=historique (where=(source_16 ne "SIFA" and source_15 ne "SIFA"));
tables natio16*natio15/nopercent norow nocol;
run;

title3 "SIFA uniquement les deux années";
Proc freq data=historique (where=(source_16 ="SIFA" and source_15 ="SIFA"));
tables natio16*natio15/nopercent norow nocol;
run;

title3 "passage de SYSCA à SIFA";
Proc freq data=historique (where=(source_16 ="SIFA" and source_15 ne "SIFA"));
tables natio16*natio15/nopercent norow nocol;
run;

***** Boursiers;
Title "15.c. Boursier au moins une fois au COLLEGE (ie SEGPA)";
Title2 "Info absente de SIFA//Non pondéré";
Proc freq data=historique;
Tables prescol*phd_qeef2*boursecol/norow nocol nopercent;
run;
Title "15.c. Boursier au moins une fois au LYCEE GENERAL/TECHNO";
Title2 "Info absente de SIFA//Non pondéré";
Proc freq data=historique;
Tables bourselyc;
Tables preslyc*phd_qeef2*bourselyc/norow nocol nopercent;
run;
```



```
***** PCS des parents;
Title "xxx. La PCS des parents est-elle dispo dans FAERE ?";
Title2 "Non pondéré";
Proc freq data=depp.generationfaere16 (where=(source="SCO "));
Tables PCS PCS2 lien1 lien2;
Tables lien1*lien2/norow nocol nofreq;
run;
Proc freq data=depp.generationfaere16 (where=(source="SIFA"));
Tables PCS PCS2 lien1 lien2 ;
run;

Proc freq data=historique ;
Tables classe16*source_16/norow nocol nopercen;
run;
```

Bibliographie

CEREQ. *Génération 2017, enquête 2020, questionnaire multimode*, 2021, 234 p.

CEREQ. *Génération 2017, Bilan méthodologique des traitements réalisés en post-enquête pour le redressement des données*, 2021, 10 p.

CEREQ. *Génération 2017, enquête 2020, dictionnaire des variables*, 2022, 137 p.

DEPP. *Fichier des apprenants (dictionnaire des variables), rentrée 2003 à 2017*, 32 p.

Jugnot S. (2021). *Les opportunités créées par la facilitation et le développement des utilisations des données administratives à des fins de recherche et de production statistique*. Céreq, Working paper, n° 9, 66 p.

Jugnot S. (2022). *Analyse du rapprochement expérimental entre le fichier historique des demandeurs d'emploi et l'enquête Génération 2004*. Céreq, Working paper, n° 11, 90 p.

Jugnot S., Vignale M. (2022). « La crise sanitaire suffit-elle à expliquer les souhaits de réorientation des jeunes ? ». *Céreq Bref*, 424, 4 p.

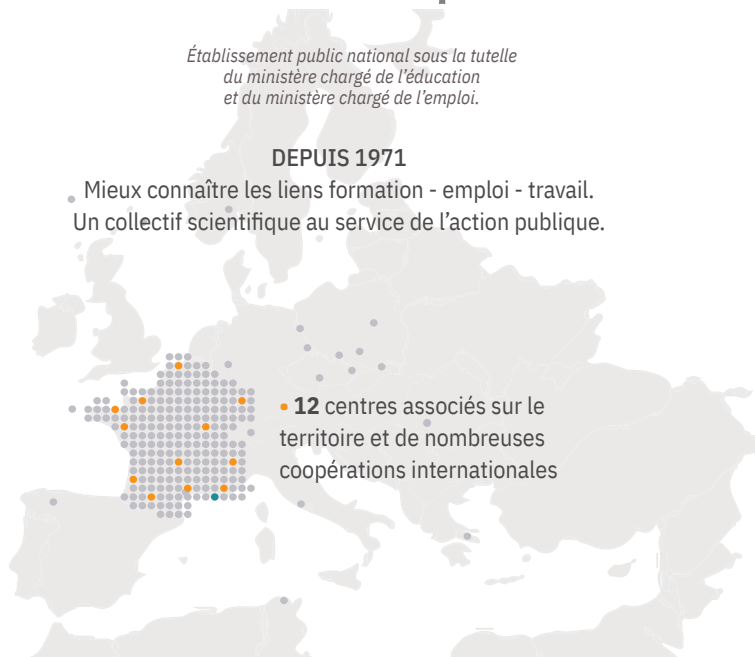
Coupié T., Gaubert E., Personnaz E. (2022). « Enquête 2020 auprès de la Génération 2017. Des parcours contrastés, une insertion plus favorable, jusqu'à... ». *Céreq Bref*, 422, 8 p.



*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



 **de 600 publications**
Accessibles librement