



HAL
open science

Que nous apprennent les fichiers SISE au sujet des inégalités de réussite dans l'enseignement supérieur ?

Guillaume Pierné

► To cite this version:

Guillaume Pierné. Que nous apprennent les fichiers SISE au sujet des inégalités de réussite dans l'enseignement supérieur ?. 2024. hal-04445677

HAL Id: hal-04445677

<https://hal.science/hal-04445677>

Preprint submitted on 8 Feb 2024

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 4.0 International License

Que nous apprennent les fichiers SISE au sujet des inégalités de réussite dans l'enseignement supérieur ?


Guillaume Pierné

2024-2 Document de Travail/ Working Paper



Economix - UMR 7235 Bâtiment Maurice Allais
Université Paris Nanterre 200, Avenue de la République
92001 Nanterre Cedex

Site Web : economix.fr
Contact : secreteriat@economix.fr
Twitter : @EconomixU



Que nous apprennent les fichiers SISE au sujet des inégalités de réussite dans l'enseignement supérieur ?

Guillaume PIERNÉ¹

2024

Résumé :

Cette étude propose une évaluation transversale des inégalités de réussite liées au sexe, à la nationalité et à la catégorie socioprofessionnelle dans l'enseignement supérieur. Elle mobilise pour cela l'exhaustivité des données contenues dans les fichiers du système d'information sur le suivi de l'étudiant pour l'année 2017-2018. Sur la base de méta-régressions, les résultats suggèrent que les inégalités de réussite sont généralisées à l'ensemble des disciplines et établissements, qu'elles sont plus fortes dans les premières années de formation et dans les filières en médecine et que leur ampleur peut varier avec la composition des formations, en particulier avec la part d'étudiants appartenant aux groupes démographiques de références.

Abstract :

This study offers a cross sectional assessment of inequalities of success related to gender, nationality and socio professional category in higher education. To this purpose, it uses the exhaustiveness of the data contained in the files of the student monitoring information system for the year 2017-2018. On the basis of meta-regressions, the results suggest that inequalities of success are generalized across subjects and educational establishments, that they are stronger in the first years of training and in medical fields and that their magnitude may vary with the demographic composition of the group of attending students , in particular with the share of students belonging to the references demographic groups

Mots clés : inégalités, réussite étudiante, enseignement supérieur, méta-analyse

Classement JEL : I23, I24

¹ Economix, Université de Paris Nanterre, g.pierne@parisnanterre.fr

Introduction

L'Enseignement supérieur joue un rôle prépondérant dans le fonctionnement des sociétés développées. Depuis la seconde moitié du 20^e siècle, l'éventail des compétences des universités françaises s'est élargi et ne se restreint plus à la production et à la diffusion de connaissances. La formation et l'insertion des futurs diplômés sur le marché du travail fait désormais partie intégrante de ces compétences et, avec près de 60% des effectifs du supérieur, les universités sont l'une des principales unités de production de futurs travailleurs qualifiés dans de nombreux secteurs. Au moyen de diplômes sanctionnant des parcours de formation et garantissant la possession de certaines aptitudes, elles permettent aux étudiants d'acquérir des signaux qui les aideront à obtenir des emplois de meilleure qualité. En 2018, le salaire horaire des diplômés de niveau bac était ainsi inférieur de 81% à celui des diplômés de niveau bac+5 (Levieil et Thélot, 2021).

A ces missions de formation et d'insertion, se conjugue l'objectif de lutte contre les inégalités de réussite éducative. Le Code de l'éducation stipule que le service public de l'éducation est organisé en fonction des étudiants et qu'il contribue à l'égalité des chances en la matière. Depuis les années 2000, plusieurs mesures visant à promouvoir cette égalité ont été instaurées par les universités (Annoot et Etienne, 2012). Dans un contexte marqué par une forte augmentation des effectifs et par une sélectivité et une mise en concurrence accrue, évaluer les inégalités de réussite dans l'enseignement supérieur, leur évolution et les déterminants de leur amplitude apparaît comme un élément important dans la mise en œuvre de mesures effectives de promotion.

Dans cette étude, nous proposons une évaluation transversale des inégalités de réussite étudiante dans l'enseignement supérieur français et de leur relation avec les caractéristiques des formations suivies, notamment leur composition. A cette fin, nous mobilisons l'exhaustivité des bases du système d'information sur le suivi de l'étudiant (SISE). Nous évaluons les différences de réussite pour trois critères d'appartenance démographique : le sexe, la nationalité et la catégorie socio-professionnelle. Au moyen de méta-régressions, nous testons notamment l'hypothèse d'une relation éventuellement non linéaire entre la part d'étudiants pouvant être rattachés au critère de référence (femme, français, enfants de cadre) et l'ampleur des inégalités. Le document est organisé

de la façon suivante. La deuxième partie présente la littérature sur les sources des inégalités de réussite à l'Université ainsi que les hypothèses testées. La troisième partie présente les données mobilisées. La quatrième partie présente les résultats. Une discussion est proposée en dernière partie.

Les sources des inégalités de réussite à l'Université

Plusieurs explications peuvent être avancées aux inégalités de réussite mesurées dans les formations universitaires (Geven et al, 2018). La première est l'existence de biais. La théorie économique avance deux explications à ces biais. L'existence d'aversion envers les membres de certains groupes démographiques (Becker, 1957) et l'imperfection de l'information concernant la performance (Arrow, 1973, Phelps, 1973). Ces explications peuvent être complétées par celles inspirées par la littérature en psychologie, comme le favoritisme endogroupe (Anderson et al., 2006) ou les biais implicites (Bertrand et al, 2005).

Des preuves empiriques de tels biais de la part d'enseignants peuvent être rapportées. En comparant les résultats obtenus à des évaluations anonymes et non anonymes, Lavy (2008) montre que les lycéens israéliens subissent des biais d'évaluation. Sprietsma (2013) montre que les enseignants des écoles allemandes notent moins bien les copies lorsqu'ils pensent qu'elles appartiennent à des élèves avec des noms à consonance turque. En Norvège, les lycéennes sont surnotées par leurs enseignants quand ces derniers sont des hommes (Falch et Naper, 2013). En Inde, les élèves des castes inférieures sont moins bien évalués, particulièrement quand leur enseignant est lui aussi de caste inférieure (Hanna et Linden, 2012). Outre les évaluations, ces biais peuvent aussi impacter l'accès à certaines formations. En Italie, les enseignants avec des stéréotypes implicites envers les immigrés sont davantage susceptibles de recommander à ces derniers des lycées et des formations de moins bonne qualité (Carlana et al, 2022). En France, Chareyron et al. (2023) montrent que les responsables de Master fournissent moins souvent des renseignements sur les modalités de candidature aux étudiants avec un nom à consonance maghrébine.

Si l'existence de biais évaluatifs dans le secteur éducatif est avérée, elle ne peut expliquer à elle seule la totalité des écarts de réussite observés dans les universités. Dans bon nombre de formations une partie des évaluations est passée de manière anonyme et des barèmes détaillés permettent de vérifier l'exactitude de la notation. La réussite ou l'échec d'un étudiant à une formation dépend généralement d'un nombre d'enseignants supérieur à celui impliqué dans le processus de sélection ou d'évaluation d'élèves ou de lycéens et l'effet d'un enseignant biaisé devrait donc y être moins préjudiciable. La composition des équipes enseignantes universitaires est en outre souvent diversifiée d'un point de vue démographique. La promotion de l'égalité des chances étant un objectif affiché par les universités, la direction prise par les biais est de plus en plus incertaine. Des biais évaluatifs visant à compenser une inégalité de représentation peuvent ainsi être identifiées dans les concours d'entrée de l'école Normale Supérieure (Bréda et Ly, 2015), dans les concours de professeurs des écoles (Bréda et Hillion, 2016) et dans de nombreux concours d'accès à la fonction publique d'état (Greenan et al, 2019).

La deuxième explication aux inégalités de réussite mesurées est la présence de traits individuels influençant la réussite universitaire corrélés avec les caractéristiques démographiques des étudiants. Si les travaux proposant d'évaluer des pratiques discriminatoires reposent sur des designs expérimentaux ou sur des expériences naturelles permettant de neutraliser l'effet de ces traits, ces derniers ne sont généralement pas totalement contrôlables dans ceux s'appuyant sur des données administratives ou d'enquêtes.

Ces traits individuels dépendent de la caractéristique démographique considérée. Pour la catégorie socioprofessionnelle, on peut par exemple supposer qu'être enfant d'ouvrier ou de parent sans profession est une proxy de caractéristiques pénalisantes dans la réussite étudiante, comme un cadre de vie moins propice aux études que celui d'un enfant de cadre. Pour le sexe, il pourrait capturer l'effet de caractéristiques déterminantes dans la réussite universitaire et distribuées en faveur des étudiantes. Fischer et al (2013) suggèrent ainsi que la réussite supérieure des lycéennes allemandes s'expliquerait par des facteurs non cognitifs tels que la motivation, la maîtrise de soi ou encore la fierté. Concernant la nationalité, l'effet mesuré pourrait représenter le différentiel de niveau entre le système éducatif français et celui du pays ou de l'aire géographique considérée.

Wang et Lin (2005) montrent que les étudiants chinois réussissent mieux que leurs homologues américains aux tests internationaux en mathématiques et que cet écart de performance est imputable aux méthodes d'enseignement et à la qualité des enseignants. La nationalité pourrait en outre capturer les différences de conditions de vie d'un étudiant de nationalité étrangère et d'un étudiant français mais aussi les différences de coût d'investissement dans la formation et d'enjeu à sa réussite.

Ces traits inobservables mais déterminants sont une source plausible d'explication des inégalités mesurées sur données non expérimentales. Si l'effet de l'appartenance démographique sur la réussite n'est alors qu'indirect, instituer l'égalité des chances implique de surmonter la différence de distribution dans ces traits individuels. Les travaux de Bourdieu et Passeron (1964) indiquent par exemple que le capital culturel des enfants des catégories favorisées est davantage valorisé par le système éducatif que celui des catégories défavorisées, ce qui favorise la reproduction sociale. Stephens et al (2012) montrent que les étudiants américains des classes défavorisées réussissent mieux les examens quand ceux-ci sont en adéquation avec leur capital culturel. De nombreux autres travaux en psychologie sociale identifient l'existence d'effets préjudiciables à certains groupes démographiques dans l'implémentation de l'évaluation de leurs capacités cognitives (Campbell, 1996, Hough et al, 2001). Pour Jaeger (2011), ces effets sont assimilables à des biais évaluatifs dès lors que les traits valorisés par les évaluations ne représentent pas des critères d'évaluation objectifs liés aux capacités des étudiants.

La dernière explication aux inégalités de réussite concerne l'effet des caractéristiques et des conditions de déroulement des formations. Les évaluations en milieu scolaire ou universitaire ne sont pas toujours indépendantes du niveau des élèves ou étudiants qui y participent et les enseignants pourraient évaluer ces derniers relativement à leurs pairs plutôt que vis-à-vis d'un standard absolu. Dans un tel cadre, l'ampleur des inégalités de réussite pourrait dépendre de la composition des classes ou formations universitaires. Une formation majoritairement composée d'étudiants appartenant à un groupe favorisé produirait davantage d'inégalités de réussite qu'une formation plutôt composée d'étudiants de groupes défavorisés et dont le niveau relatif de comparaison serait moindre. Boone et al (2018) confirment la réalité de cet effet en montrant que les enfants avec des capacités plus faibles reçoivent de meilleures

recommandations de la part de leurs enseignants lorsqu'ils sont membres de classes dont le niveau est faible. D'autres études suggèrent toutefois un effet contraire. Sur la base d'une étude expérimentale, les résultats de Kaiser et al (2017) indiquent notamment que les enseignants évalueraient avec une meilleure précision les étudiants des groupes démographiques minoritaires.

La composition des formations peut également impacter les inégalités de réussite par le biais d'effets de pairs (Barrios Fernandez, 2023). Carrel et al (2009) montrent par exemple que la performance académique d'étudiants américains en première année augmente significativement avec le niveau d'aptitudes verbales du groupe avec lequel ils passent l'essentiel de leur temps. Cet effet positif semble en outre persister tout au long du cursus de l'étudiant et être plus prononcé dans les matières scientifiques. En Italie, Bianchi (2020) montre que l'assouplissement des conditions d'accès aux grandes écoles a eu pour conséquence de diminuer le niveau des étudiants, sauf dans les matières où les nouveaux accédants étaient mieux préparés qu'en moyenne. En Hollande, Booij et al (2017) comparent les performances de groupes de travaux dirigés composés d'étudiants de même niveau avec celles de groupes composés d'étudiants de niveaux hétérogènes. Leurs résultats suggèrent que les performances des étudiants de meilleur niveau sont indépendantes de la composition de leur groupe tandis que celles des étudiants de niveau moyen et faible sont meilleures dans les groupes de niveau homogène. Cette amélioration n'est de plus pas imputable à une modification de l'enseignement. Garlick (2018) confirme le résultat concernant les étudiants de haut niveau en menant une expérience similaire sur des données sud-africaines permettant d'identifier les compagnons de chambrée. Il infirme cependant ceux concernant les étudiants de niveau moindre. De manière générale, les effets de pairs sur la réussite mesurés dans l'enseignement supérieur sont d'une ampleur bien plus modeste que dans le primaire ou le secondaire (Paloyo, 2020).

D'autres caractéristiques des formations peuvent influencer les inégalités mesurées. Les déroulements des parcours de formation induisent la mise en concurrence des étudiants à des degrés divers. Or, les étudiants de groupes démographiques distincts peuvent ressentir la concurrence de manière différente. Plusieurs travaux mettent ainsi en avant le goût plus prononcé des hommes pour la compétition et leur meilleure réussite dans les

cadres compétitifs (Gneezy et al, 2003 Niederle et Vesterlund 2007, Datta Gupta et al, 2013).

Selon les caractéristiques des formations considérées et des évaluations qui y sont menées, ces trois sources d'inégalités sont susceptibles d'expliquer une plus ou moins grande part des différentiels de réussite mesurés. Elles pourraient de plus se cumuler ou se neutraliser. La majorité des travaux cités se focalisent sur l'identification d'une seule de ces sources et, s'ils la permettent précisément, ce n'est souvent que sur un échantillon réduit. Mesurer les inégalités de manière plus transversale apparaît alors comme une démarche complémentaire importante à ces travaux. Si elle ne permet pas de distinguer précisément quelle source est à l'œuvre, elle permet l'identification des configurations dans lesquelles ces sources sont le plus susceptibles d'être présentes et de ne pas se compenser. Ces configurations peuvent par la suite faire l'objet de recherches plus précises sur les sources à l'œuvre.

Pour évaluer les inégalités de la manière la plus transversale possible, des données exhaustives sont nécessaires. En France, les bases du système d'information sur le suivi de l'étudiant (SISE) du Ministère de l'Enseignement Supérieur, de la Recherche et de l'Innovation sont celles qui répondent le plus à cette exigence d'exhaustivité. Des renseignements concernant l'ensemble des étudiants suivant une formation universitaire y sont répertoriés. Peu de travaux semblent toutefois avoir mobilisé cette source de données. Frouillou (2016) mobilise ces données dans le cadre d'une analyse de la ségrégation dans les universités franciliennes. Courtioux et al (2021) mesurent également la ségrégation au moyen des bases SISE mais dans la France entière. En comparant des indices d'exposition normalisés, ils montrent que la ségrégation est plus faible à l'université que dans le secondaire, qu'elle n'est pas forcément plus élevée dans les universités des grands ensembles urbains et qu'elle se réduit avec le niveau du diplôme considéré.

Hypothèses

Dans cette étude, outre l'évaluation transversale de l'ampleur des inégalités de réussite et de ses déterminants, nous nous focalisons plus principalement sur la relation

entre la composition des formations et l'ampleur des inégalités mesurées. L'idée derrière l'existence d'une telle relation est la suivante : dans les formations comptant peu d'étudiants des groupes démographiques de réussite supérieure, les étudiants des groupes démographiques défavorisés « fixent » le niveau de la formation et l'ampleur des inégalités est plus faible. Au fur et à mesure que la part d'étudiants des groupes démographiques de réussite supérieure augmente, cette ampleur augmente. Lorsque cette part atteint un certain point, les groupes défavorisés deviennent rares et l'ampleur des inégalités tend à se réduire. Cette réduction pourrait provenir du fait que ces groupes sont alors traités moins défavorablement pour des raisons de préservation de la diversité ou de la meilleure précision des jugements des enseignants à leur égard (Kaiser et al, 2017). Elle pourrait également être imputable à la modification des effets de pairs qui accompagne le changement dans la composition de la formation. En conséquence, la relation entre l'ampleur des inégalités et la part d'étudiants du groupe favorisé pourrait exhiber une forme en U inversé.

Plus largement, nous testons l'hypothèse d'une relation entre la composition démographique des formations et l'ampleur des inégalités. A notre connaissance, nous sommes les premiers à mobiliser l'information sur la réussite à la formation universitaire suivie contenue dans les fichiers SISE et à proposer une évaluation aussi exhaustive des inégalités de réussite dans les universités françaises. Cette évaluation contribue à la compréhension des inégalités et permet d'avancer des pistes à leur réduction.

Données

Les données sont issues des fichiers du système d'information sur le suivi de l'étudiant (SISE)². Ces fichiers recensent depuis 1995 les inscriptions et résultats aux formations suivies par les étudiants des Universités et établissements assimilés français³. Pour chaque année universitaire, deux fichiers sont produits. Le fichier relatif aux inscriptions, récolté dans le courant de l'année et le fichier relatif aux résultats, récolté dans le courant de l'année suivante. Nous mobilisons ainsi les données contenues dans le

² Ces fichiers sont disponibles sur demande auprès du Réseau Quételet

³ Les établissements d'enseignement supérieur privés, les écoles de commerce et d'ingénieur font l'objet d'enquêtes séparées et ne sont pas pris en compte dans cette étude

fichier des résultats concernant l'année universitaire 2017-2018. Chaque observation de ce fichier est relative à un étudiant, à une formation préparée et à une localisation. Un même étudiant peut donc correspondre à plusieurs observations et il en est tenu compte lorsque c'est le cas.

Plusieurs variables permettent de caractériser l'étudiant. Le sexe et la nationalité regroupée de l'étudiant, présentée en sept modalités (Afrique, Amérique, Asie, Europe, France, Non renseigné, Océanie) sont utilisées en conservant leur format initial. La catégorie socio professionnelle du parent chef de famille est renseignée en 32 modalités distinctes. Nous regroupons ces dernières en sept modalités (Agriculteurs, Artisans et Commerçants, Cadres et professions intellectuelles supérieures, Employés, Ouvriers, Professions intermédiaires, Non renseigné). Nous utilisons également l'année de naissance de l'étudiant pour calculer son âge et la série du baccalauréat ou de son équivalence (regroupée en huit modalités).

Un autre ensemble de variables permet de définir la formation suivie par l'étudiant. L'unité de rattachement pédagogique et géographique du diplôme préparé est notamment renseignée de façon précise. Le code diplôme permet de retrouver l'intitulé exact du diplôme préparé et la discipline est directement renseignée. Dans certains cas, l'observation est en outre rattachée à la préparation d'un diplôme « intermédiaire » (e.g : un DEUG dans le cadre de la préparation d'une licence) avec un code diplôme propre dont nous tenons compte.

Nous croisons l'information concernant le niveau d'études (mesuré en années d'études à partir du baccalauréat) des étudiants suivant la formation, le cycle (Licence, Master, Doctorat) dans lequel elle se situe et le niveau dans la formation préparée afin d'attribuer à chaque formation un niveau d'études de 0 (équivalence au baccalauréat) à 6 (supérieur à bac+5). Nous nous appuyons sur l'ensemble de ces caractéristiques concernant les formations pour définir les ensembles de regroupements d'étudiants suivant la même formation (formations uniques) les plus fins. Une formation unique est ainsi définie comme un regroupement d'étudiants préparant la même année du même diplôme dans la même unité de rattachement pédagogique.

Afin de déterminer la réussite des étudiants dans les formations uniques, il est nécessaire de tenir compte des spécificités des fichiers récapitulatifs des résultats. En effet, ces fichiers ne renseignent directement la réussite à la formation unique suivie par l'étudiant que lorsque cette dernière correspond à l'échéance du diplôme préparé. Ainsi, si l'on s'intéresse par exemple à la réussite dans les 3 formations uniques relatives à la préparation d'une licence, seule l'information concernant la réussite à la formation unique correspondant à la 3^e année de préparation du diplôme sera directement exploitable. Pour les formations uniques correspondant à la première et à la deuxième année, l'information ne peut être récupérée que de manière imparfaite en déterminant les caractéristiques de la formation unique suivie par l'étudiant dans le fichier des inscriptions de l'année suivante (2018-2019).

Nous procédons de la manière suivante : lorsque au moins un des résultats de la formation unique considérée est positif, nous utilisons l'information concernant ces derniers telle que fournie dans le fichier. Lorsque l'ensemble des résultats de la formation unique considérée est négatif, nous produisons un résultat redressé basé sur les caractéristiques de la formation unique suivie par l'étudiant en 2018-2019, relativement à celles de la formation unique suivie en 2017-2018. Un étudiant est alors considéré comme en réussite si le niveau d'études associé à la formation unique qu'il suit en 2018-2019 est supérieur à celui suivi en 2017-2018 et si les objets des formations uniques suivies ont un rapport. Cette définition considère donc comme réussite le passage dans une formation unique de niveau supérieur dans le cadre de la préparation d'un même diplôme mais aussi dans le cadre d'une même discipline, voire d'une discipline à l'autre si une relation peut être établie.

Lorsque l'étudiant ne peut être retrouvé dans le fichier de l'année universitaire suivante ou qu'il n'est associé qu'à une seule formation unique suivie à la fois en 2017-2018 et en 2018-2019, la détermination de la réussite redressée est directe. Lorsqu'un même étudiant suit plusieurs formations uniques, nous considérons généralement qu'il réussit à celles d'entre elles pour lesquelles il est possible de trouver une formation unique qui respecte la définition précédente et qui n'est pas déjà utilisée pour définir la réussite à

l'une des formations uniques. Sauf exceptions⁴, un étudiant suivant deux formations uniques sera donc considéré comme ayant réussi à ces formations s'il est possible de trouver au moins deux formations uniques de niveau supérieur en rapport avec ces dernières. L'édiction d'une règle générale pour ces étudiants multi inscrits est toutefois complexe et l'attribution d'un résultat pour chacune des formations qu'ils suivent doit se faire parfois quasiment au cas par cas avec une composante arbitraire. Ces observations ne représentent heureusement qu'une infime partie des données.

Cette mesure redressée de la réussite a potentiellement des conséquences sur l'évaluation des inégalités de réussite. Cela sera notamment le cas si certains groupes démographiques choisissent, indépendamment de leur réussite, de poursuivre plus ou moins fréquemment que d'autres leur parcours de formation dans le système universitaire français, la réussite étant définie en ce sens dans les formations uniques concernées. On peut par exemple penser aux étudiants étrangers qui choisiraient de ne passer qu'une année dans le système français.

Une composante importante de la réussite des étudiants concerne leur statut vis-à-vis de la formation unique suivie. Un étudiant redoublant une formation n'a sans doute pas les mêmes chances de réussite qu'un primo entrant dans le système universitaire où qu'un étudiant suivant un nouveau parcours de formation. Pour déterminer ce statut, nous mobilisons l'information contenue dans le fichier des résultats de l'année 2016-2017. A partir de cette information nous pouvons déterminer si l'étudiant était déjà présent dans le système universitaire français et, dans l'affirmative, quelles étaient les caractéristiques de la formation unique suivie. Nous répertorions notamment la localisation (même unité pédagogique, même établissement, établissement différent), la discipline (même diplôme, même discipline, autre discipline) et le niveau d'études (même année que la formation unique suivie, une année de moins, plus d'une année de moins, au moins une année de plus) de la formation unique suivie en 2016-2017 vis-à-vis de celle de 2017-2018.

⁴ Par exemple, les étudiants AJAC ou les étudiants de certaines double licence suivant deux formations uniques mais retrouvés par la suite dans une seule formation unique de niveau supérieur. Les étudiants d'une formation unique choisissant fréquemment de suivre les mêmes parcours universitaires, l'information sur la situation de la totalité des étudiants de la formation unique l'année suivante permet de s'assurer de la cohérence des choix effectués.

Au cours de cette reconstruction des données, nous mobilisons plusieurs fichiers du SISE. Tous ces fichiers ne contiennent pas toujours exactement la même information au sujet des étudiants. Il peut en effet arriver que certaines des caractéristiques individuelles ne soient pas renseignées dans les fichiers relatifs à l'année 2017-2018 mais qu'elles le soient dans les versions antérieures ou ultérieures. Lorsque c'est le cas, nous redressons les données en conséquence. Ces redressements permettent une réduction substantielle de l'information manquante.

Résultats

Statistiques descriptives

Le tableau A.1 récapitule les caractéristiques principales de l'échantillon total. Cet échantillon contient 1774471 résultats répartis dans 31264 formations uniques. L'effectif moyen de ces formations uniques est d'environ 57 étudiants et oscille entre 1 et 5737. 119 dénominations d'établissements distinctes peuvent être dénombrées, avec un effectif moyen d'environ 15000 résultats par dénomination et un nombre de formation moyenne de près de 263. Le plus petit établissement est l'Institut Physique du Globe tandis que les Universités d'Aix Marseille et de Lille rassemblent respectivement les plus grands nombre de formations uniques et d'étudiants. 75 types de diplômes sont représentés. Les types de diplômes associés aux effectifs et aux nombre de formations uniques les plus faibles sont généralement des diplômes d'Etat en médecine. Les effectifs et le nombre de formation les plus élevés peuvent être trouvés en licence et en master LMD. Selon la nomenclature la plus fine, 54 disciplines peuvent être identifiées dans l'échantillon total. Les formations uniques rattachées à la dénomination « pluridisciplinaire, droit, sciences économiques, AES » sont les moins nombreuses alors que celles associées à la dénomination « médecine » sont les plus fréquentes. En termes de résultats renseignés, ce sont les sciences juridiques qui sont associées au plus gros effectif. A l'inverse, à peine 400 résultats peuvent être liés à la discipline « langues et littérature anciennes ».

Si l'on distingue les effectifs et les formations uniques par le nombre d'années d'études après le baccalauréat, on peut voir que la majorité des résultats concerne des formations de première année (29.28% du total des résultats disponibles). En ce qui concerne les formations uniques, ces dernières sont le plus souvent de niveau supérieur à bac+5 (36.82% du total des formations uniques). A l'exception des quelques formations de niveau bac, la taille moyenne des formations uniques diminue avec le niveau. Les formations uniques regroupant le plus grand nombre de résultats sont des formations en sciences technologie santé et en sciences juridiques pour les niveaux bac+1 à bac+3, en sciences politiques et en médecine pour les niveaux supérieurs.

Le tableau A.2 décrit la composition démographique de l'échantillon total. Plus de 56% des résultats concernent des étudiantes. Pour un peu moins de 1% des résultats, le sexe de l'étudiant n'est pas renseigné. L'Université de technologie de Belfort est l'établissement avec la plus forte proportion de résultats concernant des hommes et L'ESPE de Martinique celui avec la plus faible. Les proportions d'informations manquantes ne dépassent jamais les 5% et sont nulles dans 18 établissements. La psychologie et la linguistique sont les disciplines dans laquelle on trouve le moins d'hommes parmi les étudiants, Le génie mécanique et électronique celles où la proportion d'étudiantes est la plus basse. Les formations en mathématiques appliquées en sciences sociales et en français langue étrangère sont respectivement les mieux et les moins bien renseignées au sujet de cette information.

Près de 85% des résultats concernent des étudiants de nationalité française. Dans moins de 1% des cas, la nationalité ne peut être déterminée et on peut constater que les observations concernées sont les mêmes que pour le sexe. Les étudiants étrangers sont plus nombreux dans les COMUE (Côte d'Or, Grenoble, Paris Est, Paris Saclay...), les établissements franciliens ou les instituts (Institut Physique du Globe, Institut d'Etudes Politiques, Institut National des Langues et Civilisation Orientales...) et moins nombreux dans les ESPE. La proportion d'étudiants étrangers la plus faible concerne les formations en STAPS tandis que la plus élevée concerne les formations en français langue étrangère. Si l'on distingue les étudiants étrangers par grands regroupements de nationalité, on peut voir que les étudiants africains sont les plus nombreux avec plus de 7% du total des observations renseignées. Les étudiants asiatiques et européens représentent chacun

environ 3% de ce total. Les établissements accueillant les plus fortes proportions de ces groupes de nationalités sont fréquemment localisés dans les DOM COM (Mayotte pour les étudiants africains, Guyane pour les étudiants américains, Nouvelle Calédonie pour les étudiants océaniques). Outre la discipline STAPS, les sciences de l'éducation et les formations en pharmacie accueillent moins d'étudiants étrangers. Les étudiants asiatiques, américains et européens représentent une plus forte part des résultats en français langue étrangère en littérature générale et comparée et dans les formations pluridisciplinaires en lettres et langues. Les étudiants africains en sciences générales de l'ingénieur, en électronique et en géographie.

Plus d'un résultat renseigné sur trois concerne un étudiant dont les parents sont rattachés à la catégorie socioprofessionnelle des cadres. Les résultats des étudiants avec des parents employés ou ouvriers représentent chacun environ 13% du total des observations. Globalement, l'information concernant la CSP des parents est moins bien renseignée que les autres caractéristiques. Dans 11.55% des cas, elle est indisponible. L'Université d'Avignon est l'établissement où la proportion d'étudiants dont les parents sont rattachés à la catégorie des agriculteurs est la plus forte. Ces derniers sont également plus nombreux qu'en moyenne dans les ESPE et moins nombreux en Ile de France. Les catégories des cadres et des artisans commerçants et chefs d'entreprises sont moins présentes outre-mer (Guyane, Mayotte) et plus fréquentes dans certains établissements franciliens (Paris 2, Dauphine, Polytechnique, IEP...). Les enfants de parents sans profession sont plus représentés dans les établissements d'outre-mer. Les étudiants dont les parents sont ouvriers représentent une plus grande proportion des effectifs des ESPE que des autres établissements. C'est toutefois dans les ESPE (ou dans les établissements de petite taille) que les plus fortes proportions de CSP non renseignées sont généralement identifiées. Les enfants de cadres représentent plus de la moitié des résultats en science politique et en odontologie. C'est dans les formations en administration économique et sociale que les plus fortes proportions d'ouvriers ou de sans emploi peuvent être trouvées. Les enfants d'employés représentent près d'un résultat sur cinq en langues étrangères appliquées. Les sciences religieuses et dans une moindre mesure le français langue étrangère sont les disciplines avec les plus faibles parts d'informations renseignées à ce sujet.

Au total, on peut remarquer pour l'ensemble des critères présentés une plus forte amplitude dans la composition des établissements que dans celle des disciplines, ce qui illustre probablement dans une certaine mesure la spécialisation des établissements dans ces dernières.

Afin d'étudier la réussite des étudiants dans les formations uniques, nous excluons de l'échantillon initial les formations de niveau supérieur à bac + 5 ainsi que celles de niveau bac. Il est tout d'abord complexe de déterminer les taux de réussite redressés dans les formations de ces niveaux particuliers. Les formations de niveau bac (DAEU, capacité en droit, DU en français langue étrangère...) ne représentent qu'une infime proportion des données et sortent par définition du champ d'analyse de la réussite post bac. Les formations uniques de niveau supérieur à bac + 5 (doctorat, HDR, spécialisation en médecine...) sont bien souvent « individualisées » et il n'est sans doute pas aussi judicieux de comparer la réussite de doctorants avec des sujets et des encadrants distincts que de comparer celle d'étudiants suivant un programme commun et faisant face aux mêmes modalités d'évaluation.

Les tableaux A3 et A4 décrivent les caractéristiques de l'échantillon restreint aux formations uniques de niveau bac+1 à bac+5 ainsi que sa composition démographique. L'exclusion des observations relatives aux formations uniques de niveau bac et supérieur à bac+5 conduit à la suppression de 11854 formations uniques et de 195456 résultats. En conséquence, la taille moyenne des formations uniques retenues est substantiellement supérieure à celle de l'échantillon initial. Une vingtaine de types de diplômes et trois établissements (formations sous les dénominations COMUE Paris Est, COMUE Grenoble et Institut Physique du Globe) sont exclus de l'analyse.

La restriction aux formations uniques de niveau bac+1 à bac+5 a également comme conséquence de réduire la proportion de résultats avec une information manquante au sujet du sexe de la nationalité ou de la CSP des parents. Le rapport entre la part d'hommes et de femmes demeure relativement stable alors que celui entre les étudiants étrangers et les français augmente légèrement en faveur des derniers. La proportion de résultats attribuables à des enfants de cadres diminue, celle des enfants d'ouvriers et d'employés

augmente. Les descriptions de l'échantillon total présentées préalablement demeurent valables.

Les taux de réussite mesurés globalement, dans les formations uniques à échéance et dans les formations uniques redressées sont présentés dans le tableau A.5. Cette information est distinguée selon le niveau des formations uniques et complétée par les taux de réussite par disciplines les plus faibles et élevés. Le taux de réussite global mesuré pour les femmes est de 66.98% contre 63.21% pour les hommes. Le taux de réussite des étudiants français est supérieur de près de 8.6 points de pourcentage à celui des étudiants étrangers. Les étudiants asiatiques sont ceux avec le taux de réussite le plus proche des étudiants français. Les étudiants océaniens et africains ceux avec le taux de réussite le plus faible. Les enfants d'agriculteurs sont les étudiants avec le meilleur taux de réussite globale (72.12%) suivis par les enfants de cadres (69.68%). A l'autre extrême, les enfants d'ouvriers (63.18%) et surtout de parents sans profession (53.55%) referment la marche. Les taux de réussite des résultats avec une information manquante au sujet du sexe ou de la nationalité ont un taux de réussite particulièrement élevé (plus de 90%), ce qui s'explique sans doute par leur effectif réduit. Le taux de réussite des résultats avec une CSP manquante se situe entre celui des enfants d'ouvriers et des enfants de parents sans profession.

Comme attendu, les taux de réussite redressés sont inférieurs aux taux de réussite calculés dans les formations où cette information est disponible. Le taux de réussite des femmes se réduit un peu plus que celui des hommes mais le rapport entre les taux de succès des deux groupes reste relativement similaire. Il en est de même pour le rapport entre les taux de succès des étudiants français et des étudiants étrangers. La variation du taux de réussite lors du passage d'un type de formation à l'autre n'est cependant pas la même suivant le groupe national considéré. Si la variation de celle des étudiants asiatiques est d'une ampleur comparable à celle des étudiants français, celle des étudiants africains est inférieure tandis que celles des étudiants européen, américains et océaniens est très supérieure. Les variations observées pour les taux de réussite par CSP sont elles aussi d'ampleurs différentes : elles varient moins pour les enfants d'agriculteurs et les cadres et davantage pour les enfants d'employés, de chômeurs et de sans profession.

Les taux de réussite mesurés par discipline varient considérablement et sont généralement plus bas dans les parcours pluridisciplinaires. Dans la discipline « pluridisciplinaire santé » (première année des parcours en santé), le taux de réussite mesuré est inférieur à 30%. A l'inverse, les taux de réussite en médecine, odontologie ou en génie des procédés sont supérieurs à 85%. Ces disciplines sont généralement retrouvées aux extrémités des performances quelle que soit la caractéristique considérée. Outre les potentielles différences de sélectivité entre les disciplines, des différences dans les spécificités des formations uniques associées à ces dernières (niveau, information redressée ou non, taille de l'effectifs considéré...) sont susceptibles d'expliquer ces fortes variations.

La prise en compte du niveau des formations uniques montre plusieurs choses. Premièrement, les taux de réussite redressés ne sont pas toujours inférieurs aux taux de réussite mesurés dans les formations où cette information est disponible. Cela n'est notamment pas le cas pour les formations bac+2 et bac+4. Deuxièmement la part de formations avec une information renseignée au sujet du résultat varie fortement d'un niveau à l'autre. Dans les formations de niveau bac+1 l'information est très majoritairement redressée (86.58% des formations uniques, 97.14% des résultats). Dans les formations de niveau bac+3 et bac+5 l'information est le plus souvent directement disponible (88.69% et 94.12% des formations uniques, 92.04% et 92.77% des résultats). Dans les formations de niveau bac+2 et bac+4, l'information directement disponible est majoritaire (respectivement 2/3 et 3/4 des formations et individus) mais les proportions sont plus équilibrées. Le taux de réussite le plus élevé est relatif aux formations uniques de niveau bac+3 pour lesquelles l'information concernant le résultat est disponible. Les taux les plus faibles correspondent aux formations uniques de niveau bac+1 et bac+5 pour lesquelles l'information est redressée. Les dernières représentent une faible proportion des résultats et individus. Pour les premières, la faiblesse du taux de réussite mesuré peut s'expliquer par la présence de formations uniques pouvant difficilement être suivies (e.g. : des diplômes d'université sans prolongement) mais aussi par le niveau des formations en question. Il est en effet de notoriété publique que le taux d'échec en première année est largement supérieur à celui des années suivantes.

Distinguer le niveau des formations uniques a des conséquences sur les écarts de réussite mesurés selon les caractéristiques démographiques. Les femmes ont des taux de réussite supérieurs à ceux des hommes sauf dans les formations de niveau bac+3 et bac+5 pour lesquelles l'information concernant la réussite est redressée (et qui représentent une minorité des résultats). C'est dans les formations redressées de niveau bac+2 et non redressées de niveau bac+4 que l'écart de réussite observé entre étudiant et étudiante est le plus élevé. De même, les performances des étudiants français excèdent celles des étudiants étrangers dans l'ensemble des niveaux de formation, à l'exception du niveau bac+1 quand l'information sur le résultat est disponible. C'est dans les formations uniques avec information redressée que l'écart de performance entre français et étrangers est généralement le plus élevé mais ce n'est pas le cas pour les formations de niveau bac+2. Notons que les étudiants étrangers sont par nature plus mobiles que les français, ce qui peut jouer à leur détriment dans l'évaluation des taux redressés. Un examen par groupes de nationalités suggère que cela serait particulièrement vrai pour les étudiants américains et européens : leurs taux de réussite respectifs sont particulièrement bas dans les formations à information redressée de niveau bac+3 à bac+5. Mis à part pour les rares résultats relatifs à des formations de niveau bac+1 avec information disponible, les enfants de parents sans professions et, dans une moindre mesure, d'ouvriers ont les taux de réussite les plus faibles mesurés.

Le tableau A6 présente les caractéristiques de la ou des formations suivie(s) par les étudiants en 2016. Près de 40% des résultats proviennent d'étudiants suivant le même diplôme universitaire et 50% un diplôme de la même discipline. 10% des résultats proviennent d'étudiants suivant une formation dans une discipline distincte et plus d'un tiers sont des résultats d'étudiants non présents dans le fichier 2016. De même près de 55% des résultats correspondent à des étudiants qui étaient déjà localisés dans le même établissement de diffusion parmi lesquels 85% provenaient de la même unité de rattachement pédagogique (46.35% du total). Moins de 10% des résultats proviennent d'étudiants qui suivaient une formation dans un établissement différent en 2016. Plus de 40% des résultats sont liés à des étudiants qui suivaient une formation d'exactly une année inférieure. Environ 17.5% sont liés à des étudiants suivant une formation de niveau similaire. Les parts de résultats associés à des étudiants suivant une formation de plus

d'une année inférieure (0.83%) et d'au moins une année supérieure (1.89%) sont marginales.

Dans le tableau A7, nous complétons ces statistiques en attribuant, à la manière des étudiants non présents dans le fichier 2016, un statut unique vis-à-vis de la formation en cours. Ces statuts correspondent aux croisements des caractéristiques de la formation suivie en 2016. Pour les étudiants ayant suivi plus d'une formation, nous attribuons l'ordre de priorité suivant : même diplôme, même discipline (type de formation), même unité pédagogique, même code établissement, même établissement de diffusion (localisation) même année, exactement une année avant, au moins une année après, plus d'une année avant. Outre les étudiants non présents qui représentent le statut modal, les principaux statuts attribués se rapportent aux étudiants ayant poursuivi le même diplôme dans la même unité de rattachement pédagogique (24.29%) ou ayant redoublé celui-ci dans la même unité (10.71%).

Analyse descriptive multivariée de la composition des formations uniques

Afin de mieux observer la composition des formations au niveau le plus fin, nous mobilisons une analyse en composantes principales dans laquelle chaque formation unique est définie par sa proportion d'étudiantes, d'étudiants de chacun des grands groupes de nationalités renseignés et d'enfants de chacune des catégories socio-professionnelles présentes. Pour éviter que la présence d'informations non renseignées ne capture de variabilité, nous nous restreignons dans un premier temps aux formations pour lesquelles l'information au sujet de ces caractéristiques est complète. Dans l'optique d'éviter que la valeur des proportions ne soient trop imputable à la taille des formations, nous excluons celles dont la taille est inférieure ou égale à 5 étudiants et pondérons les formations par les parts du total d'étudiants qu'elles représentent . Ces restrictions (notamment celle concernant l'information renseignée au sujet de la CSP) conduisent à l'élimination d'une part très élevée des formations uniques (près de 80%) identifiées et des résultats correspondants (plus de 90%).

Les graphiques A8 à A10 résument les résultats de cette analyse descriptive multivariée. Suivant la règle de sélection de Karlis et al. (2003), nous conservons les quatre premières

dimensions pour l'interprétation. Ces dimensions expriment plus de 50% de la variabilité totale des formations considérées en termes de composition démographique, dont plus de 20% pour la première dimension. Cette première dimension est principalement caractérisée par une opposition entre les formations uniques avec une proportion élevée d'étudiants de nationalité française et celles avec une forte proportion d'étudiants africains et, dans une moindre mesure, d'enfants de parents sans profession. A elles trois, ces proportions expriment presque 70% de l'inertie contenue sur l'axe. Sur le deuxième axe, la principale opposition concerne les formations avec une forte proportion d'enfants de cadres et celles avec une forte proportion d'enfants d'ouvriers ou d'employés (près de 80% de la variabilité exprimée). Le troisième axe oppose les formations uniques avec des proportions supérieures d'étudiantes et d'étudiants européens (et/ou américains) aux formations constituée davantage d'hommes africains. Enfin, sur le quatrième axe, on trouve une opposition entre les formations uniques avec des proportions plus élevées d'enfants de CSP moins répandues (agriculteur, artisan) où d'étudiants de nationalités moins représentées (Océanie) et les formations n'ayant pas ces caractéristiques.

A titre illustratif, nous représentons la position du taux de réussite global des formations uniques sur le cercle des corrélations et celle des disciplines sur la représentation factorielle des formations uniques. Ces positions sont respectivement calculées de la même manière que pour les proportions (taux de réussite) et sur la base des coordonnées du point moyen des formations associées à chacune des catégories (disciplines) Pour des raisons de lisibilité, nous utilisons un regroupement des disciplines en 16 modalités. Conséquence de la suppression des formations avec informations manquantes, certains de ces regroupements de disciplines sont sous (formations pluridisciplinaires) ou surreprésentés (sciences fondamentales appliqués) vis-à-vis de leurs parts initiales. Sur la base des valeurs tests (Escoffier et Pagès, 2008), nous n'identifions pas de relations significatives entre ces regroupements et l'information contenue sur les axes Le taux de réussite global des formations est mal représenté sur l'ensemble des dimensions et ne semble à première vue que peu lié aux sources de variabilité dans la composition des formations uniques qui viennent d'être évoquées.

Nous complétons cette analyse en composantes principales par une classification ascendante hiérarchique (Husson et al, 2010) des formations uniques (sur la totalité des

dimensions) avec consolidation des classes par algorithme des k means (Lloyd, 1957, Mac Queen, 1967). Sur la base de la perte d'inertie associée aux différents regroupements, nous choisissons de conserver trois classes. Le graphique A11 présente le dendrogramme tandis que le tableau A12 récapitule les caractéristiques de ces différents clusters. Nous pouvons constater que l'effectif associé à la troisième classe est largement plus restreint que celui des deux autres classes. Ces deux classes sont essentiellement caractérisées par la plus (respectivement moins) forte proportion d'enfants de cadres dans les formations uniques qui les composent. La troisième classe est quant à elle caractérisée par une proportion d'étudiants étrangers (en particulier africains) significativement plus forte. Le taux de succès global dans la première classe est supérieur de six points de pourcentage à celui de la seconde classe et de près de 22 points à celui de la troisième classe. De manière intéressante, c'est pourtant dans la première classe que le nombre de résultats redressés est le plus élevé et dans la troisième qu'il est le plus faible.

Imputation des valeurs manquantes

Pour remédier au problème des informations manquantes, nous imputons ces dernières. Pour l'information relative au sexe de l'étudiant et à sa nationalité (française ou étrangère), nous mobilisons des régressions logistiques binaires sur l'échantillon des observations pour lesquelles ces informations sont renseignées et en déduisons les probabilités prédites. Pour l'information relative aux nationalités étrangères précises des observations déterminées comme telles (ainsi que pour celles des observations pour lesquelles seul le statut français/étranger est renseigné) et aux catégories socio professionnelles, nous utilisons des régressions logistiques multinomiales⁵.

Les effets marginaux relatifs à ces estimations sont présentés dans les tableaux A13 à A15. Nous pouvons voir que les étudiants proviennent plus fréquemment d'un baccalauréat scientifique ou technologique que les étudiantes. Les étudiants suivent également plus fréquemment une formation en sciences ou en droit que les étudiantes et moins souvent

⁵ Pour les rares observations avec une nationalité étrangère précise/une CSP non renseignée et une information également manquante au sujet du baccalauréat passé, nous imputons ce dernier de la même manière. Nous remplaçons également l'année de naissance des rares observations pour lesquelles elle n'est pas renseignée par l'année de naissance moyenne de l'échantillon sur lequel l'estimation est effectuée.

en langues ou en santé. La probabilité de trouver un étudiant dans une formation de niveau supérieur à bac + 1 est également significativement plus faible. Le fait d'avoir réussi sa formation apparaît comme un prédicteur significatif du sexe de l'étudiant. La dichotomie français étranger suggère que les étudiants étrangers proviennent moins fréquemment de baccalauréats en sciences économiques et sociales ou technologiques hors tertiaire que les étudiants français. Ces étudiants sont également plus nombreux dans les formations de niveau supérieur à bac+1 et moins nombreux en STAPS ou en sciences humaines et sociales. Ici également, la réussite à la formation apparaît comme un prédicteur significatif, mais dans une moindre ampleur que pour le sexe.

Les régressions multinomiales montrent que les étudiants africains sont moins fréquemment titulaires d'un baccalauréat scientifique que leurs homologues asiatiques ou européens. Ils suivent également plus souvent une formation en sciences naturelles ou en AES que les autres étudiants étrangers et, à l'exception des étudiants américains, sont plus fréquemment dans des formations de niveau bac+4 ou bac+5 que dans des niveaux inférieurs. Le fait d'être de nationalité africaine est enfin associé à un taux de réussite significativement inférieur à celui des étudiants américains asiatiques ou européens. Concernant la CSP, nous pouvons voir que les étudiants enfants de cadres sont généralement plus souvent titulaires d'un baccalauréat scientifique et moins souvent d'un autre type de baccalauréat. Les enfants de cadres sont en outre plus souvent en médecine ou en lettres et moins souvent en langue ou en STAPS. Le niveau de la formation influence la probabilité d'appartenance aux catégories socioprofessionnelles avec plus d'enfants de cadres et moins d'enfants d'employés ou d'ouvriers dans les formations bac+4 ou bac+5 que dans les bac+1. Comparativement aux enfants de cadres, la réussite à la formation suivie est associée à une probabilité d'appartenance supérieure pour les enfants d'agriculteurs et de parents de professions intermédiaires, inférieure pour les catégories employés et ouvriers.

Ces modèles permettent-ils de prédire les valeurs manquantes de manière adéquate ? Pour l'information concernant le sexe, nous choisissons un seuil de classification de 0.4. Ce seuil permet d'améliorer la sensibilité du modèle (63% contre 43% avec un seuil de 0.5) pour un taux de prédictions erronées similaire. Le taux de faux positif augmente toutefois de façon importante (34% contre 14%). Appliqué à l'échantillon des valeurs

manquantes, le modèle prédit ainsi qu'environ 60% des valeurs manquantes comme femmes (plus de 80% avec un seuil de 0.5). Le modèle de prédiction de la nationalité (française/étrangère) semble également convenablement performant (sensibilité de 72%, taux de prédictions erronées inférieur à 5%) mais ses performances ne sont pas impactées par la modification du seuil de classification et le modèle prédit la quasi-totalité des valeurs manquantes comme françaises. Les modèles multinomiaux ont tendance à surreprésenter les étudiants africains (80% des modalités manquantes prédites) et les enfants de cadres (85% des modalités manquantes). Les informations manquantes concernant la nationalité et le sexe ne représentent qu'une infime (moins de 1%) proportion des observations et l'imputation de ces informations ne devrait pas impacter les résultats. L'imputation de la catégorie socioprofessionnelle (environ 10% des observations) peut toutefois avoir des conséquences plus importantes et les résultats doivent ainsi être interprétés avec précaution.

Les graphiques A16 à A18 récapitulent les résultats de cette analyse en composantes principales. Sur la base de la même règle de sélection qu'énoncée précédemment, nous conservons les quatre premières dimensions. Ces dimensions représentent plus de 55% de la variabilité totale des formations en termes de composition démographique, dont un peu moins de la moitié pour la seule première dimension. La principale opposition sur ce premier axe ne concerne plus les proportions d'étudiants d'origine française et africaine mais les proportions d'étudiants d'origine française d'un côté et celles d'enfants de cadres de l'autre. Dans une moindre mesure, les proportions d'étudiants asiatiques et d'enfants de professions intermédiaires ou d'employés captent également une part de la variabilité contenue sur l'axe. Sur la deuxième dimension, les principales contributions proviennent de la proportion d'enfants de cadres et d'étudiants de nationalités africaines d'un côté et, de l'autre des proportions d'ouvriers et de chômeurs (plus de 70% de la variabilité contenue sur l'axe). Sur la troisième dimension, la principale opposition identifiable concerne les formations avec des proportions plus élevées d'étudiants de nationalités africaines et celles qui sont plus féminines ou associées à des proportions plus fortes d'étudiants de nationalité européenne ou américaine (80% de la variabilité contenue sur l'axe). La quatrième dimension illustre enfin une opposition entre les formations composées d'étudiants de nationalités (océaniques) où dont les

parents occupent des professions (professions intermédiaires, artisans, agriculteurs) moins représentées dans l'échantillon.

Comme lors de l'analyse sur le sous échantillon des formations avec information complète, le taux de réussite global des formations est plutôt mal représenté sur l'ensemble des dimensions. Concernant les regroupements de disciplines, nous identifions au moyen des valeurs tests plusieurs relations significatives avec les axes. Les regroupements « sciences humaines et sociales » et « STAPS » sont ainsi liés significativement (à 10%) au premier axe. Ceci signifie qu'ils sont caractérisés par des proportions d'étudiants de nationalité française supérieures et d'enfants de cadres inférieures. Les regroupements « AES », « langues » et « médecine » peuvent quant à eux être significativement associés au second axe et aux oppositions qui y sont illustrées. Pour le troisième axe, on retrouve les langues, l'art et les sciences humaines et sociales d'une part (associations positives et significatives) et d'autre part les sciences appliquées et les sciences économiques (associations négative et significatives). Pour le quatrième axe, on peut identifier des associations significatives avec les sciences appliquées (positives) et la médecine (négatives), ce qui suggère que ces regroupements de disciplines sont les plus (resp. les moins) divers en termes de nationalités/catégories socioprofessionnelles les moins fréquentes.

Le graphique A19 et le tableau A20 récapitulent la classification ascendante hiérarchique avec consolidation des classes par k-means. Nous retenons trois classes. La première, plus nombreuse (près de 3/4 des observations, plus de 2/3 des formations uniques) et homogène, est caractérisée par une proportion supérieure d'étudiants de nationalité française et par des proportions plus faibles d'étudiants africains, asiatiques ou européens. En termes de catégorie socioprofessionnelle, la proportion d'enfants de cadres est plus faible tandis que celles d'enfants de professions intermédiaires, d'employés et d'ouvriers sont légèrement supérieures. La seconde classe (près d'1/4 des résultats et 30% des formations) est plus masculine (différence significative à 10.4%) et est essentiellement caractérisée par sa forte proportion d'étudiants africains et d'enfants de cadre. La troisième classe, la plus petite et hétérogène, est plus féminine et est composée très majoritairement d'étudiants américains, asiatiques et européens. Le taux de réussite moyen est relativement proche de celui de l'échantillon global des formations

étudiées dans la première classe, supérieur de 5 points de pourcentage dans la seconde classe et inférieur de 15 points dans la troisième classe. La part de résultats redressés est toutefois plus forte dans la première classe que dans les deuxième et troisième classes, ce qui pourrait influencer les différences mesurées.

Mesure des inégalités de réussite

Afin d'évaluer les inégalités de réussite dans l'enseignement supérieur, une première possibilité est de mobiliser l'échantillon global tout en tenant compte des caractéristiques relatives au parcours et à la formation des étudiants. Le tableau A21 présente les effets marginaux relatifs à des régressions logistiques de la réussite étudiante dans lesquelles les écarts types estimés sont clustérisés à l'échelle des formations uniques. La première colonne introduit uniquement le sexe de l'étudiant, sa nationalité et la catégorie socioprofessionnelle de ses parents en tant que prédicteurs. La seconde colonne tient également compte du baccalauréat passé, du niveau de la formation suivie, de la situation de l'étudiant l'année précédente, de son âge vis-à-vis du reste de la formation, du nombre de formations suivies et du fait que le résultat attribué à l'étudiant soit redressé ou non. La troisième colonne ajoute la discipline de la formation suivie (sur 16 niveaux) et la dernière colonne l'établissement. Pour limiter le nombre de modalités associées à cette dernière variable, nous regroupons les ESPE et les COMUE avec les universités qui leur correspondent (passage de 116 à 69 modalités).

Les résultats montrent que les étudiants ont une probabilité de réussite inférieure de près de 6 points de pourcentage à celui des étudiantes lorsque les caractéristiques énumérées précédemment sont contrôlées. Les étudiants africains (- 6.6 points), européens (- 2.8 points) et américains (-1.8 point) réussissent moins bien que les étudiants français tandis que les étudiants asiatiques réussissent légèrement mieux que ces derniers. Comparativement aux enfants de cadres et à l'exception des enfants d'agriculteurs qui réussissent le mieux, les enfants des autres catégories socioprofessionnelles ont des taux de réussite significativement inférieurs, avec des différences particulièrement marquées pour les catégories sans profession (-7.6 points) et ouvrier (-3.3 points). Les étudiants titulaires d'un bac S réussissent significativement mieux que les étudiants titulaires d'un bac ES. A l'inverse, les étudiants titulaires d'un bac professionnel ou technologique

réussissent significativement moins bien. Comme suggéré par les statistiques descriptives, la réussite est plus faible en première et en deuxième année. La situation de l'étudiant l'année précédente influe également le taux de réussite. Par rapport à un nouvel entrant, les étudiants qui repassent une formation unique de même niveau, ceux qui passent à un niveau immédiatement supérieur et ceux qui reprennent à un niveau inférieur à celui passé réussissent significativement mieux. Les étudiants dont l'âge est différent de l'âge modal de la formation unique ont des taux de réussite significativement inférieurs. Le taux de réussite est enfin plus faible dans les formations uniques avec résultat redressé, quand l'étudiant suit plusieurs formations uniques ou quand la formation unique est assurée dans un ESPE.

Le degré de contrôle de la régression influence l'ampleur des effets marginaux estimés, mais pas de manière uniforme. Pour la nationalité et la catégorie socioprofessionnelle, les effets mesurés sont plus forts quand ils ne sont pas contrôlés. C'est le contraire pour le sexe, ce qui suggère que les caractéristiques dont nous tenons compte sont davantage en la faveur des étudiants en matière de réussite que ne le sont celles des étudiantes. Ces contrôles permettent d'améliorer le pouvoir explicatif de la régression bien que celui-ci reste modeste.

Si ces premiers résultats confirment l'existence d'inégalités de réussite globale, ils peuvent être complétés par une évaluation plus fine. Ainsi, une seconde possibilité consiste à mesurer ces inégalités par discipline et par établissement. Nous effectuons à cette fin des régressions logistiques dans chaque discipline et établissement⁶. Ces régressions tiennent compte des mêmes caractéristiques que la régression menée sur l'échantillon global. Les régressions par établissement contrôlent de la discipline et celles par discipline contrôlent de l'établissement. Dans de rares cas, il n'est pas possible de tenir compte de la spécification complète du modèle et celui-ci doit être simplifié. Il n'est parfois enfin simplement pas possible d'estimer l'effet des caractéristiques démographiques d'intérêt pour des raisons d'effectifs.

⁶ Pour les disciplines, nous utilisons le regroupement en 54 modalités. Pour les établissements, nous considérons les ESPE séparément et ne regroupons que les formations uniques assurées dans les COMUE (111 modalités)

Une dernière possibilité consiste à mesurer les inégalités de réussite dans chacune des formations uniques définies. Pour que cela soit possible, les formations uniques évaluées doivent pouvoir être associées à la fois à des étudiants qui réussissent et à des étudiants qui échouent. Afin de ne pas estimer d'effets sur des effectifs trop réduits, nous excluons de l'analyse les formations uniques dont l'effectif est inférieur ou égal à 5 étudiants. Le nombre de formations uniques étant important, nous conduisons l'estimation de la façon suivante. Dans un premier temps nous estimons pour chacune des formations uniques des régressions logistiques où les 3 critères d'inégalités d'intérêt sont introduits de manière séparée ou simultanément. Les résultats obtenus dans cette première étape nous permettent de déterminer quels effets du sexe de la nationalité et de la catégorie socioprofessionnelle peuvent être évalués dans chacune des formations uniques. Dans un second temps, nous ajoutons aux critères d'inégalités d'intérêt les contrôles pouvant être introduits à ce niveau d'analyse : l'âge vis-à-vis de la formation, la série du baccalauréat, le nombre de formations uniques suivies et les caractéristiques de la formation suivie en 2016-2017. Pour ce dernier contrôle, nous utilisons une version simplifiée (sur 9 positions) des statuts uniques vis-à-vis de la formation en cours présentés dans le tableau A.7. Finalement, nous conservons les versions les plus contrôlées possible de nos régressions dès lors que les effets marginaux et écarts types robustes à l'hétéroscédasticité sont estimés de façon convenable. 12885 formations uniques sont ainsi associées à l'estimation de l'effet d'au moins un des critères d'inégalités d'intérêt et plus de 90% de l'effectif des résultats observés de bac+1 à bac+5 est analysé. Bien sûr, tous les critères ne peuvent être estimés avec la même fréquence. Le sexe est le critère le plus fréquemment estimé avec 11331 estimations. Pour la nationalité africaine, plus de 6000 estimations sont réalisées. Les autres nationalités donnent lieu à un nombre d'estimation moindre. Pour la catégorie socioprofessionnelle, les estimations concernant les enfants d'ouvriers et d'employés sont plus fréquentes tandis que celles concernant les enfants d'agriculteurs sont moins nombreuses.

Les tableaux A.22 et A.23 récapitulent la distribution des effets marginaux estimés à chaque niveau d'analyse. Dans le tableau A.22 nous distinguons les effets positifs et négatifs au seuil de significativité conventionnel, tandis que dans le tableau A.23, nous opérons cette distinction à un seuil de 10%. A ce seuil, des inégalités significatives de réussite en défaveur des hommes peuvent être identifiées dans environ 80% des

disciplines et établissements et dans près de 20% des formations uniques. Certains ESPE et les universités technologiques semblent davantage épargnés par les inégalités de réussite liées au sexe. Il en est de même dans certaines matières plutôt littéraires mais aussi en physique. Dans les sciences religieuses et les parcours pluridisciplinaires de santé, l'inégalité de réussite est tournée en faveur des hommes.

Les étudiants africains sont également concernés par des inégalités de réussite en leur défaveur dans près de 70% des disciplines, 60% des établissements et 25% des formations uniques. On peut toutefois noter que ces étudiants réussissent mieux que les français dans les formations aux métiers de l'ingénieur et de manière équivalente dans certaines disciplines scientifiques (mathématiques et informatique, mathématiques appliquées aux sciences sociales...) médicales (odontologie, pharmacie...) et certains parcours pluridisciplinaires (droit et sciences politiques, langues, sciences...). Les étudiants asiatiques et européens réussissent mieux en psychologie et en langues.

Sans surprise, les enfants de parents sans profession et dans une moindre mesure d'ouvriers sont ceux pour qui des inégalités de réussite en leur défaveur peuvent être le plus fréquemment identifiées. Pour les premiers cités, une réussite significativement inférieure à celle des cadres est détectée dans près de 80% des disciplines et 70% des établissements. C'est dans des établissements de type ESPE que la pénalité semble être la moins importante. Des inégalités de réussite en faveur de ces groupes peuvent également être identifiées, notamment au niveau d'analyse des formations uniques, mais bien moins fréquemment qu'en leur défaveur.

Méta analyse des résultats

Ces évaluations plus fines des inégalités de réussite associées aux critères considérés nous permettent de construire une méta-base où chaque observation est définie par les caractéristiques observables de la formation unique à laquelle elle se rapporte et par les estimations qui ont pu y être réalisées. A partir de cette méta-base, nous pouvons mener une méta-analyse à effets aléatoires (Borenstein et al, 2010) des résultats obtenus et améliorer la fiabilité des effets mesurés. Nous considérons que chaque effet marginal estimé est distribué selon une loi normale dont l'espérance est égale

à la vraie valeur de l'effet et dont la variance est définie comme la somme de la variance de l'effet marginal estimé (intra-études) et de la variance inter études⁷. L'effet agrégé est alors calculé comme une moyenne des effets marginaux pondérée par l'inverse de cette somme. Sa variance est estimée avec une modification à la Knapp et Hartung (2003).

Nous obtenons ainsi un effet agrégé pour chacun des critères d'inégalités. Le tableau A.24 récapitule ces effets agrégés dans le cadre de méta analyses conduites sur chaque niveau d'observation considéré. Plusieurs choses peuvent être notées. Premièrement, les effets agrégés estimés dans les méta analyses sur les établissements et les disciplines sont très proches et vont globalement dans le même sens que ceux indiqués par la régression sur l'ensemble des données. Ce n'est toutefois pas le cas pour l'effet de la nationalité asiatique et de la CSP professions intermédiaires qui cessent d'être significatifs. Cela pourrait s'expliquer par le fait que les étudiants caractérisés par ces critères font partie de disciplines/d'établissements associés à de meilleurs (resp. moins bons) taux de réussite qu'en moyenne mais qu'ils ne réussissent pas mieux (resp. moins bien) que les étudiants français/enfants de cadres au sein de ces disciplines/établissements. Deuxièmement, les effets agrégés mesurés ont parfois tendance à être plus négatifs quand la méta analyse est menée sur les formations uniques. En effet si les effets agrégés associés au sexe, à la nationalité européenne ou à la CSP des ouvriers demeurent relativement stables quel que soit le niveau d'analyse, ceux associés aux autres nationalités et CSP semblent davantage en défaveur des étudiants concernés. Les effets agrégés de la nationalité asiatique et de la CSP agriculteurs deviennent notamment négatifs et significatifs. Ces différences dans les effets mesurés pourraient être attribuables à des effets de composition dans les formations uniques pouvant être retenues dans la méta analyse correspondante. Ces dernières constitueraient alors un échantillon davantage défavorable aux étudiants concernés comparativement à leur situation générale dans l'enseignement supérieur.

Une question importante concerne l'effet du redressement du résultat dans les formations uniques où l'information concernant la réussite n'est pas disponible. Cette dernière étant alors définie comme le fait d'être retrouvé dans le fichier SISE de l'année suivante dans une formation unique de niveau supérieur en relation avec la formation unique suivie, il

⁷ Nous utilisons la méthode du maximum de vraisemblance résiduel (Patterson et Thompson, 1971) pour estimer cette composante.

pourrait être possible que les inégalités qui y sont mesurées soient plus ou moins intenses. Ce sera notamment le cas si les étudiants associés à certains critères quittent plus ou moins fréquemment le champ d'observation de SISE, indépendamment du fait qu'ils aient réussi ou pas leur formation. Le tableau A.25 présente la distribution des effets mesurés dans les formations uniques où l'information concernant le résultat est disponible et l'effet agrégé correspondant. Globalement, les inégalités mesurées y sont plus défavorables pour les étudiants de nationalité étrangère et pour les enfants d'agriculteurs et plus favorables pour les hommes et les autres catégories socio professionnelles. Ces différences ne sont toutefois pas de nature à modifier les constats présentés.

Afin de mieux tenir compte de cet effet mais aussi d'observer comment les inégalités mesurées sont influencées par les caractéristiques des formations, nous mobilisons la méthode de la méta régression (Berkey et al, 1995). Nous nous restreignons à l'étude de huit modalités de nos critères d'inégalité : le sexe, les regroupements de nationalités africaines, asiatiques et européennes, les enfants d'agriculteurs, d'employés, d'ouvriers et de parents sans profession. Nous expliquons l'ampleur des inégalités mesurées au sein des formations uniques par un ensemble de contrôles comprenant le statut de la formation vis-à-vis du redressement du résultat, le logarithme de l'effectif de la formation unique, le taux de réussite global de la formation unique, le nombre d'années d'études après le baccalauréat de la formation unique, le type de l'établissement dans lequel se trouve la formation (ESPE ou non), la discipline à laquelle se rattache la formation unique. Pour cette dernière, nous procédons à un regroupement en 4 grands ensembles : sciences humaines et sociales (SHS), sciences, littéraire et médecine. Outre ces contrôles, nous tenons compte de la proportion d'étudiants du critère de référence (femme, français, cadre) et du carré de cette proportion afin de tester l'hypothèse d'une relation éventuellement non linéaire entre l'ampleur des inégalités et la composition de la formation unique.

Les tableaux A.26 et A.27 détaillent les estimations obtenues au moyen de ces méta régressions. Deux spécifications sont proposées. Dans la première, l'effet brut du contrôle

est présenté⁸. Dans la seconde, les contrôles sont introduits simultanément. Nous retrouvons les effets du redressement du résultat mentionnés dans les spécifications brutes mais seules les estimations concernant le regroupement de nationalités africaines et les enfants d'agriculteurs semblent impactées quand les autres modérateurs sont pris en compte. L'accroissement de l'effectif des formations semble réduire les inégalités mesurées. A l'inverse, à l'exception de celles en défaveur des hommes, elles augmentent avec le taux de réussite global de la formation. Les estimations concernant les formations uniques de niveau bac+1 sont plus défavorables aux hommes et aux enfants de non cadres, en particulier d'ouvriers et de parents sans professions, que celles concernant des niveaux supérieurs. Le fait que la formation soit dispensée en ESPE n'influence pas l'ampleur des effets estimés pour le sexe et les nationalités mais diminue celle associée aux effets concernant les enfants d'agriculteurs, d'employés ou d'ouvriers. Par rapport aux SHS, les inégalités mesurées dans les formations du regroupement de disciplines plutôt littéraires semblent être moins défavorables aux étudiants des regroupements de nationalités asiatiques et européennes et aux enfants d'employés, d'ouvriers et de parents sans professions. Globalement, les inégalités mesurées dans les formations plutôt scientifiques sont de même ampleur que celles mesurées en SHS, à l'exception de celles concernant les étudiants de nationalités africaines qui sont moins fortes. Les inégalités mesurées dans les formations en médecine sont plus fortes qu'en SHS pour les étudiants de nationalités européennes et asiatiques ainsi que pour les enfants de non cadres. Elles sont moins fortes pour les hommes.

Les effets associés à la part d'étudiants du critère de référence montrent que cette dernière influence négativement et de manière relativement constante l'ampleur des inégalités de réussite envers les hommes. Nous n'identifions pas de relation entre la proportion d'étudiants de nationalité française et les inégalités auxquelles font face les étudiants associés aux regroupements de nationalités étudiés. Un effet négatif et non linéaire peut être identifié pour les étudiants de nationalités africaines mais seulement dans la spécification brute. Des relations significatives en U inversé entre les inégalités mesurées et la part d'étudiants cadre dans la formation peuvent être identifiées pour les

⁸ Pour des raisons de présentation, les estimations de la première spécification sont présentées dans la même colonne. Les constantes associées à chacun des modèles n'y sont pas reportées. Elles sont très majoritairement significatives et de même signe que les effets agrégés préalablement mesurés.

enfants d'employés, d'ouvriers et de parents sans profession. Ces relations sont robustes à l'inclusion des autres contrôles. Le changement de signe est évalué comme se produisant aux environs de la moitié de l'effectif de la formation (43.47% pour les enfants de parents sans profession, 44.87% pour ceux d'employés, 51.29% pour ceux d'ouvriers).

Afin de compléter notre analyse de la relation entre inégalités de réussite et composition des formations, nous reprenons la répartition des formations suggérée par la classification ascendante hiérarchique et l'incluons dans des méta régressions contrôlées de la même façon. Les résultats présentés dans les tableaux A28 et A29 suggèrent que les inégalités de réussite sont moins défavorables aux hommes dans les classes 2 et 3 caractérisées par des compositions respectivement plus masculines et féminines que celles de la classe 1. Pour les étudiants de nationalité étrangère, les classes 2 et 3, composées de moins d'étudiants français, semblent moins défavorables en matière d'inégalités. L'ampleur des inégalités de réussite associées à la catégorie socio professionnelle des parents semble par contre relativement indépendante de la classe d'appartenance de la formation. Seules les inégalités de réussite subies par les enfants d'employés semblent plus élevées dans la classe 3. Le fait que les classes d'appartenance 2 et 3 soient définies simultanément par leur composition plus fortes d'étudiants étrangers et de cadres pourrait expliquer ce résultat. La valeur et la significativité des autres contrôles introduits dans les méta régressions ne sont que peu modifiées dans cette spécification. L'effet positif de la taille de l'effectif de la formation cesse d'être significatif pour les hommes et pour les enfants d'ouvriers tandis que l'effet négatif du taux de réussite global devient significatif pour l'ensemble des CSP et cesse d'être significatif pour les étudiants africains. Les effets associés au niveau de la formation sont globalement équivalents. L'effet positif des ESPE disparaît pour les enfants d'agriculteurs et devient négatif et significatif pour les étudiants africains. Les différences d'ampleur des inégalités par regroupement de disciplines restent les mêmes à l'exception des hommes pour qui elles sont davantage marquées dans cette spécification.

Discussion

Cette évaluation de l'ampleur des inégalités et de sa relation avec les caractéristiques des formations suivies nous enseigne plusieurs choses. Premièrement,

les inégalités de réussite étudiante semblent être généralisées à la majorité des disciplines et établissements de l'enseignement supérieur français. Deuxièmement, ces inégalités semblent plus fortes dans les premiers niveaux de formation de l'enseignement supérieur. Ce résultat pourrait s'expliquer par l'effet des inégalités de réussite dans les premiers niveaux de formation sur les caractéristiques moyennes des groupes démographiques comparés aux groupes de référence dans les niveaux supérieurs. Troisièmement, si la majorité des disciplines semblent concernées par les inégalités de réussite, toutes ne le sont pas dans les mêmes mesures. Les formations en médecine, réputées pour leur forte sélectivité et mise en concurrence, semblent notamment être particulièrement inégalitaires. Les formations plutôt littéraires, qui pourraient donner davantage de poids à des éléments subjectifs dans l'évaluation que les formations plus scientifiques, apparaissent comme plus favorables aux enfants des CSP défavorisées et aux étudiants des nationalités européennes et asiatiques. Le différentiel concernant ces derniers pourrait s'expliquer par leurs meilleures performances dans certaines disciplines rattachées au regroupement, notamment les langues étrangères où il peuvent posséder un avantage vis-à-vis des étudiants français. Pour les autres, la différence pourrait provenir de cette variation dans le poids accordé à la subjectivité dans l'évaluation. Les enseignants de ces disciplines auraient alors tendance à mettre cette subjectivité au service des groupes défavorisés. Il est toutefois également possible que ce résultat, comme les autres concernant les regroupements de disciplines, proviennent d'effets de composition. Les enfants de cadres suivant des formations littéraires formeraient alors un groupe moins différent des autres en termes de caractéristiques induisant la réussite que les enfants de cadres suivant des formations dans d'autres disciplines. Les formations plutôt scientifiques sont relativement proches en matière d'inégalités des formations en SHS. Elles semblent cependant moins inégalitaires pour les étudiants de nationalités africaines qui les suivent pour une forte part.

L'hypothèse d'une relation entre la composition de la formation et l'ampleur des inégalités semble partiellement vérifiée. Nos résultats suggèrent une relation curvilinéaire entre l'ampleur des inégalités de réussite des CSP défavorisées et la proportion d'enfants de cadres et une relation linéaire entre les inégalités de réussite des hommes et la proportion de femmes. Nos résultats suggèrent également que les formations composées davantage ou de moins de femmes que ce qui est généralement

observé sont moins inégalitaires pour les hommes et que celles composées de moins (resp. plus) d'étudiants de nationalité française (resp. de cadres) sont moins (resp. plus) inégalitaires pour les étudiants étrangers (enfants d'employés). Les indications apportées par les deux spécifications testées ne sont toutefois pas identiques. L'absence de relation entre la composition des formations et les inégalités de réussite attribuables à la catégorie socioprofessionnelle dans la seconde spécification pourrait s'expliquer par les caractéristiques des classes 2 et 3, composées à la fois davantage d'étudiants étrangers et d'enfants de cadres. Les résultats concernant les inégalités de réussite attribuables à la nationalité suggèrent l'existence d'un effet de seuil dans la composition des formations plutôt qu'un effet continu. Pour les hommes, la première spécification suggère un effet linéaire tandis que la seconde suggère des inégalités de réussite moindres à la fois dans les formations comportant moins et davantage de femmes. Ici encore les caractéristiques des classes, notamment la troisième, pourraient expliquer la différence de résultats.

Ces résultats suggèrent essentiellement trois pistes en matière de recommandation dans la lutte contre les inégalités de réussite dans l'enseignement supérieur. La première est une focalisation accrue sur les formations suivies dans les premières années du parcours universitaire et pour lesquelles les inégalités de réussite semblent plus fortes. La seconde est une attention particulière aux formations en médecine. Si la grande majorité des disciplines sont concernées par les inégalités de réussite, ces dernières y apparaissent comme plus prégnantes, en particulier pour les non cadres et les étudiants étrangers. La troisième piste concerne la composition des groupes de travail au sein des formations uniques. Dans les formations avec les effectifs les plus élevés, les étudiants sont répartis en groupe de travaux dirigés et passent les épreuves de contrôle continu en leur sein, avec des évaluations pouvant être spécifiques au groupe. Nos résultats suggèrent que des groupes moins mixtes en termes de sexe, nationalité ou catégorie socioprofessionnelle pourraient produire des inégalités de réussite de moindre ampleur. Cette mesure pourrait être expérimentée relativement facilement et sans engendrer de coûts par les responsables de formation.

Cette étude comporte plusieurs limites. La première est conséquence des données mobilisées. Si elles sont relativement exhaustives, elles ne permettent pas de connaître précisément la cause imputable aux inégalités de réussite et in fine à leur variation.

Comme indiqué, il apparaît peu probable que les inégalités de réussite soient entièrement imputables à des biais évaluatifs. Les inobservables corrélés aux groupes démographiques et ayant pu impacter les différences mesurées ne peuvent toutefois pas être précisément identifiés et sont d'autant plus nombreux que les fichiers SISE n'ont pas été spécifiquement conçus pour pratiquer ce genre d'évaluation. De même, la raison avancée à la variation de l'ampleur des inégalités de réussite avec la composition de la formation reste hypothétique et, si la relation identifiée la conforte au moins partiellement, davantage de preuves empiriques sont nécessaires pour une vérification formelle. Enfin, la construction des fichiers SISE et le fait que les résultats concernant la réussite à la formation ne soient pas toujours directement disponibles complique la mise en œuvre d'une évaluation exhaustive. Nous avons choisi de conserver les évaluations des formations concernées dans la mesure où nous l'avons contrôlé dans la production des résultats et où celles-ci ne sont pas toujours différentes de celles mesurées dans les formations où l'information est disponible. Lorsque c'est le cas, les résultats sont plus favorables aux critères plus défavorisés alors que la crainte était plutôt que ce soit le contraire pour des raisons indépendantes de la réussite, notamment pour les étudiants étrangers. En ce sens, les résultats sont plutôt rassurants concernant l'impact de ce redressement de l'information. Rendre l'information concernant la réussite disponible pour chaque niveau de formation plutôt que pour le cursus total du diplôme préparé permettrait toutefois aux chercheurs un accès facilité à cette dernière.

Une deuxième limite porte sur la répartition des formations uniques induite par la classification ascendante hiérarchique. Cette répartition en groupes de tailles très inégales et avec, pour les plus petits, des caractéristiques de composition assez particulières, rend l'interprétation des résultats concernant la classe d'appartenance compliquée. Le dendrogramme montre qu'une répartition en un nombre supérieur de classes conduirait avant tout à éclater les classes de taille la plus restreinte, ce qui ne semble pas souhaitable. Les résultats concernant les classes d'appartenance sont ainsi à interpréter avec une précaution particulière.

Une troisième limite concerne la période analysée. Si cette évaluation est transversale et exhaustive, elle ne porte cependant que sur les résultats d'une année universitaire. Si l'information de l'année précédente et de l'année suivante est mobilisée, c'est seulement

en guise de contrôle ou de redressement. La réplication de ce type d'études sur d'autres années permettrait d'évaluer comment les inégalités et les relations identifiées évoluent dans le temps. Il serait alors possible de constituer une base en panel et de tenir compte de l'hétérogénéité individuelle des formations. La création d'une telle base apparaît comme un prolongement possible de cette étude qui constitue néanmoins un premier point de comparaison.

Conclusion

Cette étude propose une évaluation transversale des inégalités de réussite dans l'enseignement supérieur français et de leur relation avec les caractéristiques des formations suivies. Pour cela, elle mobilise l'exhaustivité des données des fichiers SISE relatifs à l'année 2017-2018 et analyse au moyen de méta régressions les inégalités de réussite associées au sexe, à la nationalité et à la catégorie socioprofessionnelle. A notre connaissance, c'est la première fois qu'une étude aussi exhaustive des inégalités de réussite dans l'enseignement supérieur français est menée.

Les résultats de cette étude suggèrent que ces inégalités de réussite sont largement généralisées à l'ensemble des disciplines et des établissements. Elles apparaissent toutefois comme plus importantes dans les premières années de l'enseignement supérieur et dans les filières en médecine. Les résultats suggèrent également que l'ampleur des inégalités pourrait dépendre de la composition des formations, notamment de la part d'étudiants associés aux critères de référence, suivant des relations linéaires, curvilinéaires ou de seuil. Ainsi moduler la composition des sous-groupes d'étudiants composant les formations suivies dans les premières années semble une piste intéressante de lutte contre ces inégalités.

Références

- Anderson, Lisa R., Roland G. Fryer, and Charles A. Holt. (2006). Discrimination: Experimental Evidence from Psychology and Economics. In *Handbook on the Economics of Discrimination*, edited by William M. Rodgers, III, 97–115. Northampton, MA: Edward Elgar
- Annot, E. and Étienne, R. (2012) , “Égalité des chances, universités et territoires, trois éléments enchevêtrés dans le paysage des réformes de l’enseignement supérieur”, *Éducation et socialisation*
- Arrow, Kenneth J. (1973). The Theory of Discrimination. In *Discrimination in Labor Markets*, edited by Orley Ashenfelter and Albert Rees, 3–33. Princeton, NJ: Princeton University Press.
- Barrios-Fernandez, A. (2023, October 18). Peer Effects in Education. *Oxford Research Encyclopedia of Economics and Finance*.
- Becker, Gary. 1957. *The Economics of Discrimination*. Chicago: University of Chicago Press
- Berkey CS, Hoaglin DC, Mosteller F, Colditz GA.b (1995) A random-effects regression model for meta-analysis. *Stat Med*. Feb 28;14(4):395-411.
- Bertrand, M., Chugh, D., and Mullainathan, S. (2005). Implicit discrimination. *American Economic Review*, pages 94–98
- Bianchi, N. (2020). The Indirect Effects of Educational Expansions: Evidence from a Large Enrollment Increase in University Majors. *Journal of Labor Economics* 38 (3), 767–804
- Booij, A. S., E. Leuven, and H. Oosterbeek (2017). Ability Peer Effects in University: Evidence from a Randomized Experiment. *The review of economic studies* 84 (2), 547–578
- Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research synthesis methods*, 1(2), 97-111.
- Bourdieu, Pierre, and Jean-Claude Passeron. (1964). *Les Héritiers: Les étudiants et la culture*. Paris: Les Editions de Minuit.
- Bréda, Thomas, and Son Thierry Ly. (2015). Professors in Core Science Are Not Always Biased against Women: Evidence from France. *American Economic Journal: Applied Economics* 7(4): 53–75.
- Bréda, Thomas, and Mélina Hillion. (2016). Teaching Accreditation Examinations Reveal Grading Biases Favor Women in Male-Dominated Disciplines in France. *Science* 353(6298): 474–8.
- Boone, S., Thys, S., Van Avermaet, P., & Van Houtte, M. (2018). Class composition as a frame of reference for teachers? The influence of class context on teacher recommendations. *British Educational Research Journal*, 44(2), 274–293
- Carrell, S. E., R. L. Fullerton, and J. E. West (2009). Does your Cohort Matter? Measuring Peer Effects in College Achievement. *Journal of Labor Economics* 27 (3), 439–464.
- Campbell, John P. (1996). Group Differences and Personnel Decisions: Validity, Fairness, and Affirmative Action. *Journal of Vocational Behavior* 49(2): 122–58.
- Carlana, Michela, Eliana La Ferrara and Paolo Pinotti. (2022). "Implicit Stereotypes in Teachers' Track Recommendations." *AEA Papers and Proceedings*, 112 : 409-14.

- Chareyron, S., Erb, L.-A., & L'Horty, Y. (2023). ASSESSING DISCRIMINATION IN ACCESS TO HIGHER EDUCATION: RESULTS FROM A FIELD EXPERIMENT. *Annals of Economics and Statistics*, 151, 121–145.
- Pierre Courtioux, Tristan-Pierre Maury, Johan Seux. (2021) Ségrégation sociale à l'Université : des disparités académiques sur la période 2006-2016. halshs-03167787
- Datta, Gupta, Nabanita Anders Poulsen, and Marie-Claire Villeval. (2013). Gender Matching and Competitiveness: Experimental Evidence. *Economic Inquiry* 51(1): 816–35.
- Escoffier, Jérôme et Brigitte Pagès. (2008). *Analyses factorielles simples et multiples: objectifs méthodes et interprétation*. Paris: Dunod, 4ème éd.
- Falch, Torberg, and Linn R. Naper. (2013). Educational Evaluation Schemes and Gender Gaps in Student Achievement. *Economics of Education Review* 36: 12–25
- Fischer, F., Schult, J. & Hell, B. (2013). Sex differences in secondary school success: why female students perform better. *Eur J Psychol Educ* 28, 529–543
- Frouillou L (2016) Les mécanismes d'une ségrégation universitaire francilienne : carte universitaire et sens du placement étudiant ». Thèse de géographie. Université Paris I Pantheon-Sorbonne.
- Garlick, R. (2018). Academic Peer Effects with Different Group Assignment Policies: Residential Tracking versus Random Assignment. *American Economic Journal: Applied Economics* 10 (3), 345–69
- Geven, S., Batruch, A., & van de Werfhorst, H. (2018). Inequality in teacher judgements, expectations and track recommendations: a review study. Amsterdam: University of Amsterdam. <https://zoek.officielebekendmakingen.nl/blg-864911>
- Gneezy, U., Niederle, M.I, et Rustichini, A.(2003). "Performance in competitive environments: Gender differences", *Quarterly Journal of Economics*, 118(3), pp. 1049–1074
- Greenan, N., Lanfranchi, J., L'Horty, Y., Narcy, M. and Pierné, G. (2019), Do Competitive Examinations Promote Diversity in Civil Service?. *Public Admin Rev*, 79: 370-382. <https://doi.org/10.1111/puar.13053>
- Hanna, Rema N., and Leigh L. Linden. (2012). Discrimination in Grading. *American Economic Journal: Economic Policy* 4(4): 146–68
- Hough, Leatta M., Frederick L. Oswald, and Robert E. Ployhart. (2001). Determinants, Detection and Amelioration of Adverse Impact in Personnel Selection Procedures: Issues, Evidence and Lessons Learned. *International Journal of Selection and Assessment* 9(1–2): 152–94
- Husson, F., Josse, J. & Pagès J. (2010). Principal component methods - hierarchical clustering - partitionnal clustering: why would we need to choose for visualizing data?. *Rapport technique*.
- Jæger, M. M. (2011). Does cultural capital really affect academic achievement? New evidence from combined sibling and panel data. *Sociology of Education*, 84(4), 281–298
- Kaiser, J., Südkamp, A., & Möller, J. (2017). The effects of student characteristics on teachers' judgment accuracy: Disentangling ethnicity, minority status, and achievement. *Journal of Educational Psychology*, 109(6), 871-888.

- Karlis, Dimitris & Saporta, Gilbert & Spinakis, Antonis. (2003). A Simple Rule for the Selection of Principal Components. COMMUNICATIONS IN STATISTICS Theory and Methods Vol. 32. No. 3. 643-666. 10.1081/STA-120018556.
- Knapp, G., and J. Hartung. (2003). Improved tests for a random effects meta-regression with a single covariate. *Statistics in Medicine* 22: 2693-271
- Lavy, Victor. (2008). Do Gender Stereotypes Reduce Girls' or Boys' Human Capital Outcomes? Evidence from a Natural Experiment. *Journal of Public Economics* 92(10-11): 2083-105
- Levieil, A .et Thélot, H., (2021) INSEE Focus n° 248, Septembre
- Lloyd, S. P. (1957). Least squares quantization in PCM. Technical Report RR-5497, Bell Lab, September
- MacQueen, J. B. (1967). Some methods for classification and analysis of multivariate observations. In L. M. Le Cam & J. Neyman (Eds.), *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability* (Vol. 1, pp. 281-297). California: University of California Press.
- Niederle M. et Vesterlund L. (2007). « Do Women Shy Away from Competition? Do Men Compete Too Much? », *Quarterly Journal of Economics*, 122(3),1067-1101.
- Paloyo, A. (2020), Chapter 21 - Peer effects in education: recent empirical evidence, Editor(s): Steve Bradley, Colin Green, *The Economics of Education (Second Edition)*, Academic Press, Pages 291-305,
- Patterson, H. Desmond, and Robin Thompson. (1971). Recovery of Inter-block Information When Block Sizes Are Unequal. *Biometrika* 58(3): 545-54.
- Phelps, E. (1972). The statistical theory of racism and sexism. *The American Economic Review*, 62(4):659-61
- Stephens, N. M., Fryberg, S. A., Markus, H. R., Johnson, C. S., & Covarrubias, R. (2012). Unseen disadvantage: How American universities' focus on independence undermines the academic performance of first-generation college students. *Journal of Personality and Social Psychology*, 102(6), 1178-1197
- Wang, J., & Lin, E. (2005). Comparative Studies on U.S. and Chinese Mathematics Learning and the Implications for Standards-Based Mathematics Teaching Reform. *Educational Researcher*, 34(5), 3-13

ANNEXE

Tableau A.1 Caractéristiques de l'échantillon total

1) : en nombre d'individus (2) : en nombre de formations uniques. Les écarts types sont présentés entre parenthèse. Le nombre de formations uniques est présenté entre crochets dans la seconde partie du tableau

	N	Effectif		
		Min	\bar{X}	Max
Nombre d'individus	1774471			
Nombre de formations uniques	31264	1	56.75 (137.69)	5737
Nombre d'établissements (1)	119	13	14911.52 (15906.06)	69678
Nombre d'établissements (2)	119	1	262.72 (278.01)	1262
Nombre de diplômes (1)	75	1	23659.61 (94026.03)	753505
Nombre de diplômes (2)	75	1	416.85 (1162.92)	6844
Nombre de disciplines (1)	54	399	32860.57 (39739.07)	186324
Nombre de disciplines (2)	54	13	578.96 (1041.52)	7223
Formation bac : +0	19433 [343]	1	56.66 (95.53)	1209
Formation bac : +1	519479 [2921]	1	177.84 (306.56)	5737
Formation bac : +2	293412 [2930]	1	100.14 (139.72)	1817
Formation bac : +3	299856 [4555]	1	65.83 (108.19)	1634
Formation bac : +4	226519 [3950]	1	57.34 (98.32)	2557
Formation bac : +5	239749 [5054]	1	47.44 (94.6)	2479
Formation bac : > +5	176023 [11511]	1	15.29 (67.41)	3280

Tableau A.2 Composition de l'échantillon total

En pourcentage du total

Caractéristique	Total	Etablissement		Discipline	
		Min	Max	Min	Max
Sexe :					
Homme	42.97	15.94	81.13	19.28	85.50
Femme	56.07	17.92	83.88	13.64	80.32
Non renseigné	0.96	0.00	5.00	0.03	8.06
Nationalité :					
Français	84.55	48.87	99.60	26.74	96.66
Etranger	14.48	0.41	47.96	3.17	65.20
<i>Dont</i>					
Afrique	7.18	0.00	28.95	1.63	26.77
Amérique	1.18	0.00	20.44	0.23	12.20
Asie	2.82	0.00	15.38	0.26	33.75
Europe	3.22	0.00	17.85	1.02	17.59
Océanie	0.05	0.00	13.73	0.00	0.73
Non renseigné	0.96	0.00	5.00	0.03	8.06
CSP des parents					
Agriculteur	1.63	0.00	7.48	0.29	3.54
Artisan	8.23	0.00	16.98	3.66	11.06
Cadre	35.19	7.36	63.49	14.72	55.00
Chômeur	6.71	0.00	41.51	2.77	15.48
Employé	13.10	0.00	23.65	3.14	19.29
Intermédiaire	10.29	0.00	14.29	4.01	14.80
Ouvrier	13.31	0.00	27.69	6.20	20.28
Non renseigné	11.55	0.00	84.62	2.52	58.01

*

Tableau A.3 Caractéristiques de l'échantillon des formations bac+1 à bac+5

1) : en nombre d'individus (2) : en nombre de formations uniques. Les écarts types sont présentés entre parenthèse

	N	Effectif		
		Min	\bar{X}	Max
Nombre d'individus	1579015			
Nombre de formations uniques	19410	1	81.35 (161.42)	5737
Nombre d'établissements (1)	116	33	13612.20 (13894.33)	62430
Nombre d'établissements (2)	116	3	167.33 (169.14)	872
Nombre de diplômes (1)	54	1	29241.02 (109599.8)	753505
Nombre de diplômes (2)	54	1	359.44 (1142.54)	6844
Nombre de disciplines (1)	54	116	29241.02 (34402.57)	177579
Nombre de disciplines (2)	54	13	359.44 (434.05)	2375
Nombre de formation bac : +1	519479 [2921]	1	177.84 (306.56)	5737
Nombre de formation bac : +2	293412 [2930]	1	100.14 (139.72)	1817
Nombre de formation bac : +3	299856 [4555]	1	65.83 (108.19)	1634
Nombre de formation bac : +4	226519 [3950]	1	57.34 (98.32)	2557
Nombre de formation bac : +5	239749 [5054]	1	47.44 (94.6)	2479

Tableau A.4 Composition de l'échantillon des formations bac+1 à bac+5

En pourcentage du total

Caractéristique	Total	Etablissement		Discipline	
		Min	Max	Min	Max
Sexe :					
Homme	42.95	15.94	81.02	18.50	87.55
Femme	56.34	18.02	83.88	12.09	81.36
Non renseigné	0.71	0.00	48.48	0.00	8.29
Nationalité :					
Français	85.63	36.36	99.60	29.68	96.92
Etranger	13.66	0.40	35.15	2.95	62.02
<i>dont</i>					
Afrique	7.10	0.00	28.94	1.50	26.82
Amérique	1.07	0.00	20.84	0.15	11.75
Asie	2.46	0.00	14.42	0.20	31.22
Europe	2.98	0.00	17.03	0.97	17.83
Océanie	0.03	0.00	6.10	0.00	0.79
Non renseigné	0.71	0.00	48.48	0.00	8.29
CSP des parents					
Agriculteur	1.63	0.24	7.46	0.00	3.44
Artisan	8.44	1.86	23.81	3.73	11.06
Cadre	34.41	7.36	67.25	14.33	56.33
Chômeur	7.02	0.00	41.51	2.95	15.48
Employé	13.99	0.00	25.51	3.13	18.85
Intermédiaire	10.77	1.57	14.39	4.31	14.90
Ouvrier	13.91	0.00	27.69	6.08	20.28
Non renseigné	9.83	0.00	66.67	2.51	59.84

Tableau A.5 Taux de réussite des formations bac+1 à bac+5

En pourcentage du total. (1) : formations pour lesquelles le résultat concernant la réussite est directement disponible. (2) : formations pour lesquelles le résultat concernant la réussite est redressé.

Caractéristique	Total			Discipline		Bac+1		Bac+2		Bac+3		Bac+4		Bac+5	
	T	(1)	(2)	Min	Max	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Global	65.53	74.99	53.78	29.02	90.80	64.9	47.45	68.53	70.42	80.66	75.13	68.92	73.06	78.85	48.21
Sexe :															
Homme	63.21	72.38	52.49	30.13	88.07	58.85	45.75	66.94	65.37	77.34	79.58	65.47	76.20	77.09	51.48
Femme	66.98	76.44	54.89	28.47	93.01	62.17	48.86	69.78	73.92	82.70	68.71	70.89	70.71	79.52	46.17
Non renseigné	91.54	96.99	0.16	11.11	100	96.69	0.36	93.31	0.00	97.72	0.00	94.59	0.00	97.48	0.00
Nationalité :															
Français	66.53	76.75	54.78	29.62	91.13	59.89	48.15	70.12	71.87	82.80	81.27	71.42	77.43	80.15	50.42
Etranger	57.96	64.37	45.66	20.41	82.73	62.88	40.85	53.52	59.08	65.27	48.18	58.19	55.20	72.57	21.70
<i>dont</i>															
Afrique	56.04	60.33	46.46	16.33	83.51	50.67	39.08	49.62	54.93	60.00	76.44	53.36	66.98	70.60	22.20
Amérique	57.16	66.73	39.56	0.00	100	69.33	38.36	52.97	60.99	66.96	19.01	63.84	39.45	71.57	12.35
Asie	63.66	70.04	50.68	0.00	90.70	72.47	44.66	56.15	63.05	73.34	56.92	63.63	66.97	76.03	20.09
Europe	57.90	69.12	43.04	16.94	88.89	42.03	42.19	60.95	64.26	71.42	18.22	67.15	38.41	75.23	23.67
Océanie	53.31	69.89	33.91	0.00	100	71.43	43.57	47.06	61.54	73.96	4.76	80.00	14.55	75.64	0.00
Non renseigné	91.54	96.99	0.16	11.11	100	96.69	0.36	93.31	0.00	97.72	0.00	94.59	0.00	97.48	0.00
CSP parents															
Agriculteur	72.12	78.95	61.81	31.26	92.70	62.03	55.79	74.60	74.40	85.17	79.72	73.32	79.90	79.55	53.48
Artisan	66.05	75.66	54.29	26.96	93.31	61.16	47.59	69.14	71.18	81.05	73.65	69.22	75.58	80.60	49.75
Cadre	69.68	76.66	61.48	35.73	91.59	57.96	54.68	69.62	72.74	82.08	81.14	71.02	76.84	81.04	60.42
Chômeur	53.55	67.71	38.87	16.26	86.21	57.44	33.35	60.44	61.16	74.12	70.11	61.73	69.01	72.33	33.60
Employé	64.34	76.78	51.82	23.89	90.62	62.6	46.09	69.88	72.74	82.32	75.96	72.13	78.57	82.02	46.21
Intermédiaire	67.88	78.46	56.41	27.86	90.42	56.44	50.14	72.76	74.14	84.09	78.20	72.61	76.30	82.89	56.86
Ouvrier	63.18	73.92	49.31	21.72	92.68	58.09	43.31	68.42	70.39	79.87	76.53	67.30	79.60	77.60	39.84
Non renseigné	60.52	70.05	34.66	21.45	87.37	70.74	34.75	56.59	45.21	75.22	34.56	62.14	35.64	73.41	6.92
N	1579015					2.86	97.14	65.29	34.71	92.04	7.96	75.19	24.81	92.77	7.23
Formations	19410					13.42	86.58	68.19	31.81	88.69	11.31	75.97	24.03	94.12	5.18

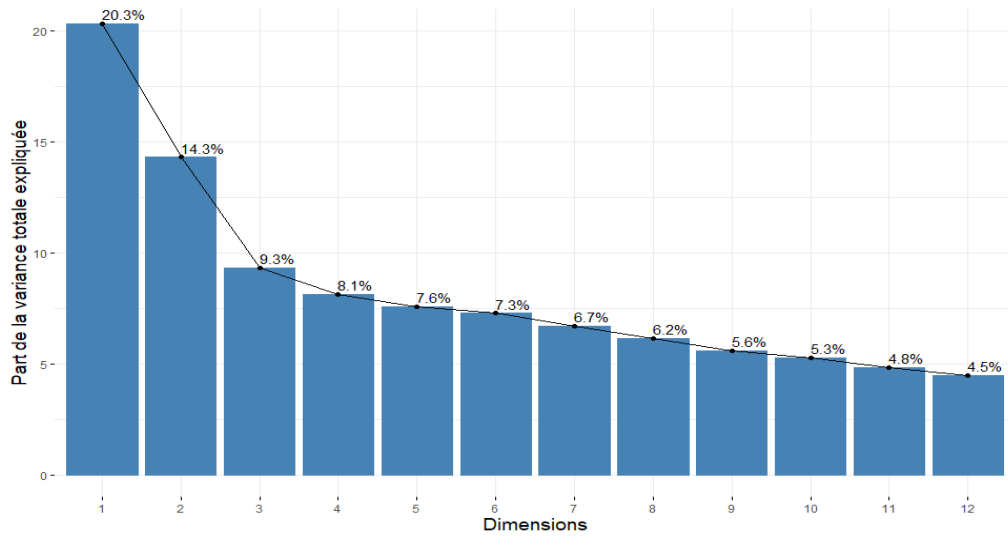
Tableau A.6 Caractéristiques des formations suivies par les étudiants en 2016-2017

Type de formation	Pourcentage total
Même diplôme	38.33
Même discipline	53.28
Autre discipline	10.72
Localisation	
Même unité pédagogique	46.35
Même code établissement	51.38
Même établissement diffusion	54.28
Etablissement différent	9.16
Niveau par rapport à la formation en cours	
Plus d'un an avant	0.83
Un an avant	43.52
Même année	17.53
Au moins un an après	1.89

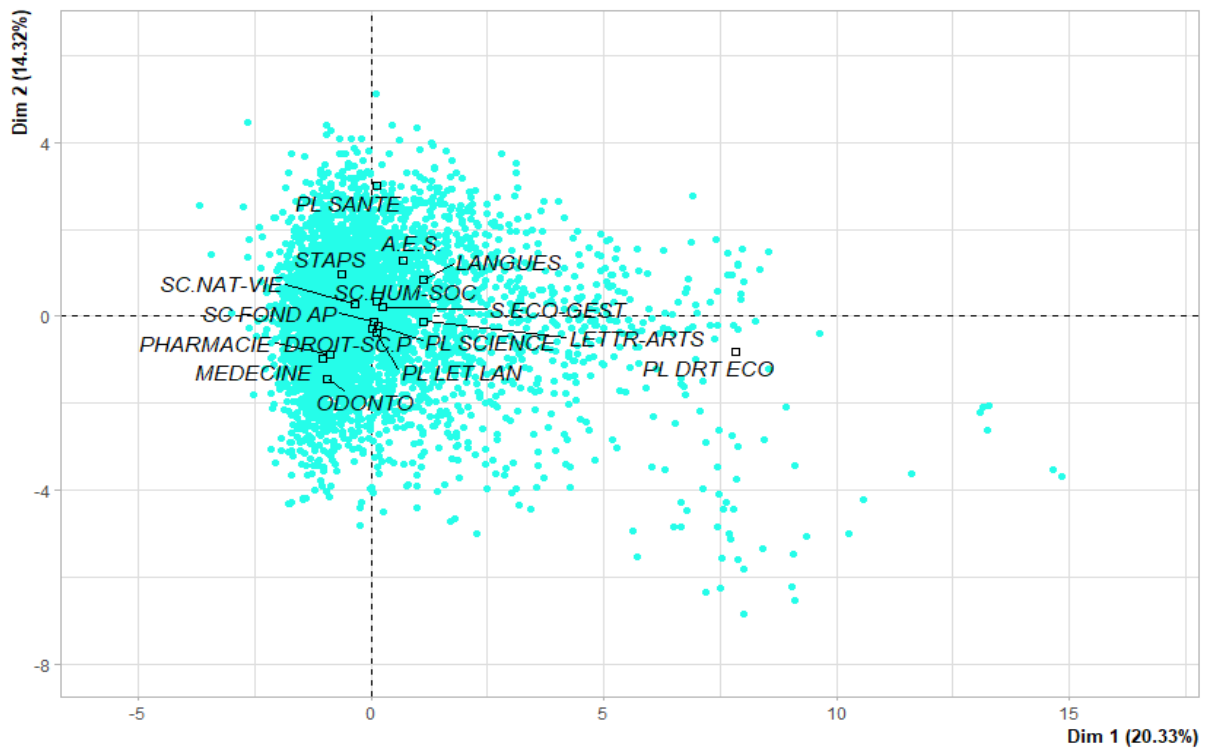
Tableau A.7 Caractéristiques principales des formations suivies par les étudiants en 2016-2017

Formation suivie :		Niveau de la formation suivie par rapport au niveau de la formation en cours			
		Plus d'un an avant	Un an avant	Même année	Au moins un an après
Même unité pédagogique	Même diplôme	0.03	24.29	10.71	0.08
	Même discipline	0.08	7.39	0.98	0.28
	Autre discipline	0.15	1.77	0.40	0.18
Même code établissement	Même diplôme	0.00	0.42	0.12	0.00
	Même discipline	0.02	1.23	0.32	0.08
	Autre discipline	0.09	1.20	1.26	0.30
Même établissement diffusion	Même diplôme	0.00	1.25	0.53	0.00
	Même discipline	0.00	0.30	0.29	0.01
	Autre discipline	0.01	0.24	0.21	0.04
Etablissement diffusion différent	Même diplôme	0.00	0.59	0.29	0.01
	Même discipline	0.06	2.91	0.73	0.28
	Autre discipline	0.12	1.79	1.55	0.47
Total présents		0.56	43.38	17.39	1.73
Non présent		36.92			

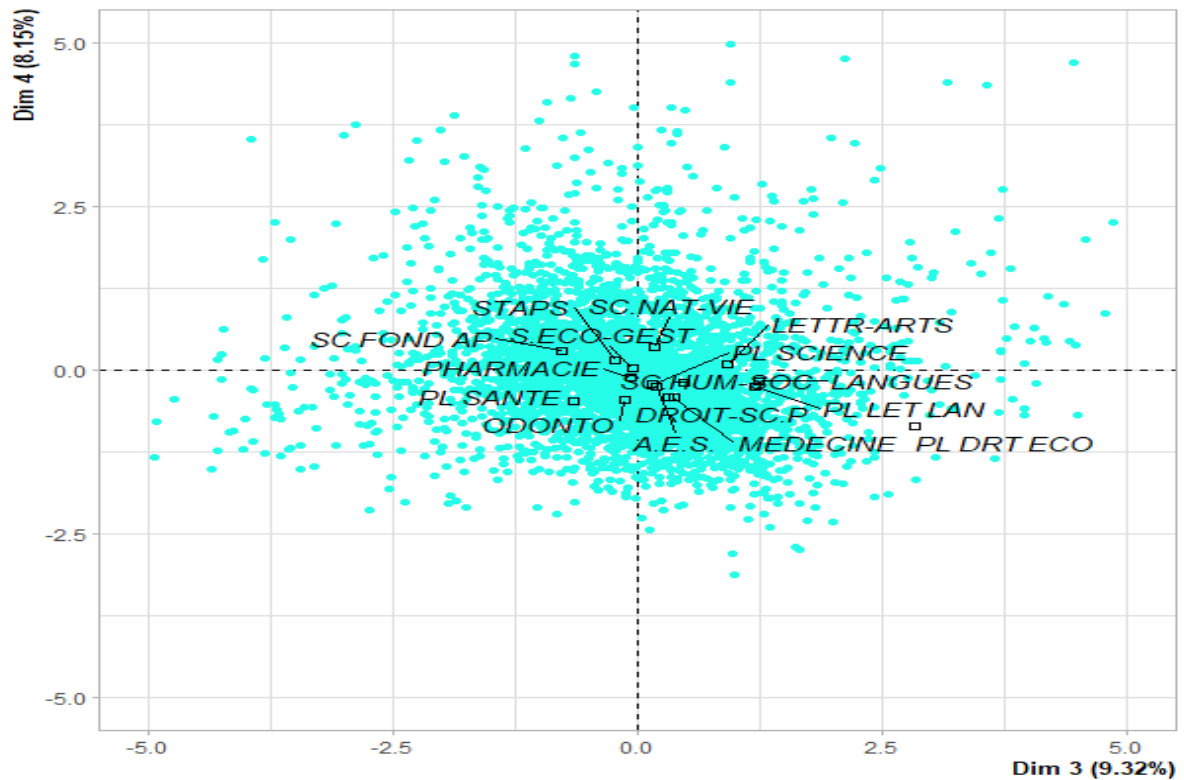
Graphique A.8 Distribution des valeurs propres – Échantillon des formations uniques avec information complète



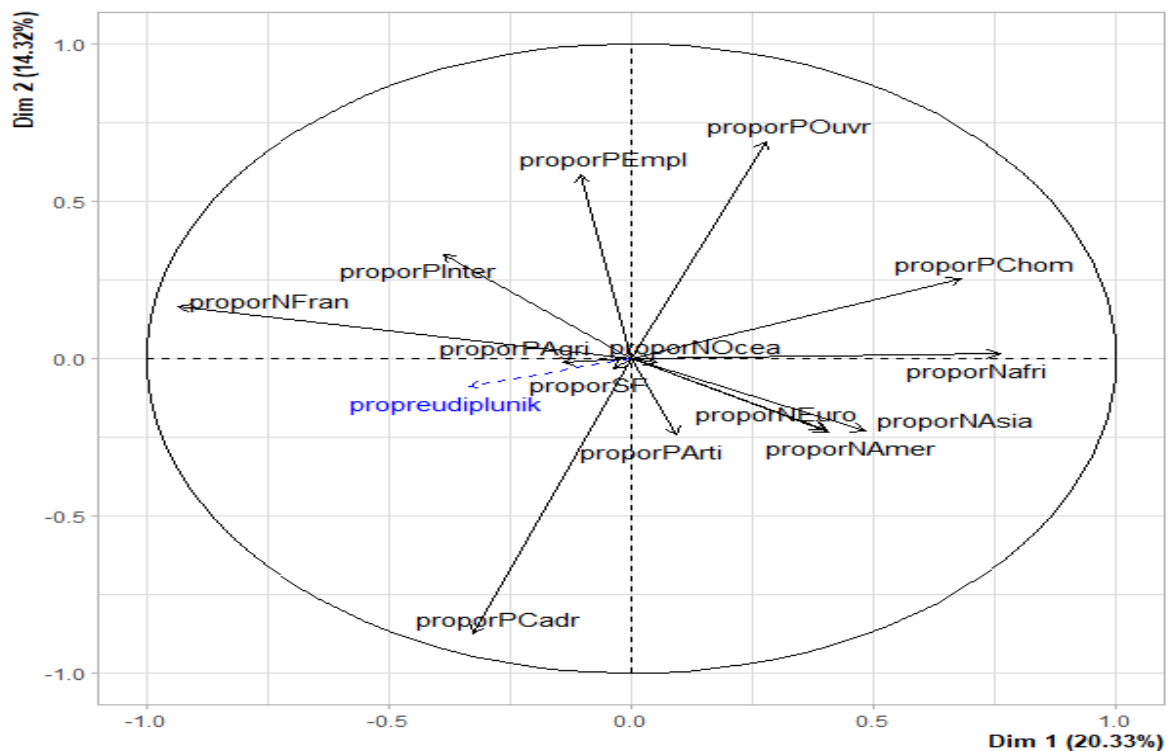
Graphique A.9 (1/2) Représentation factorielle des individus – Échantillon des formations uniques avec information complète



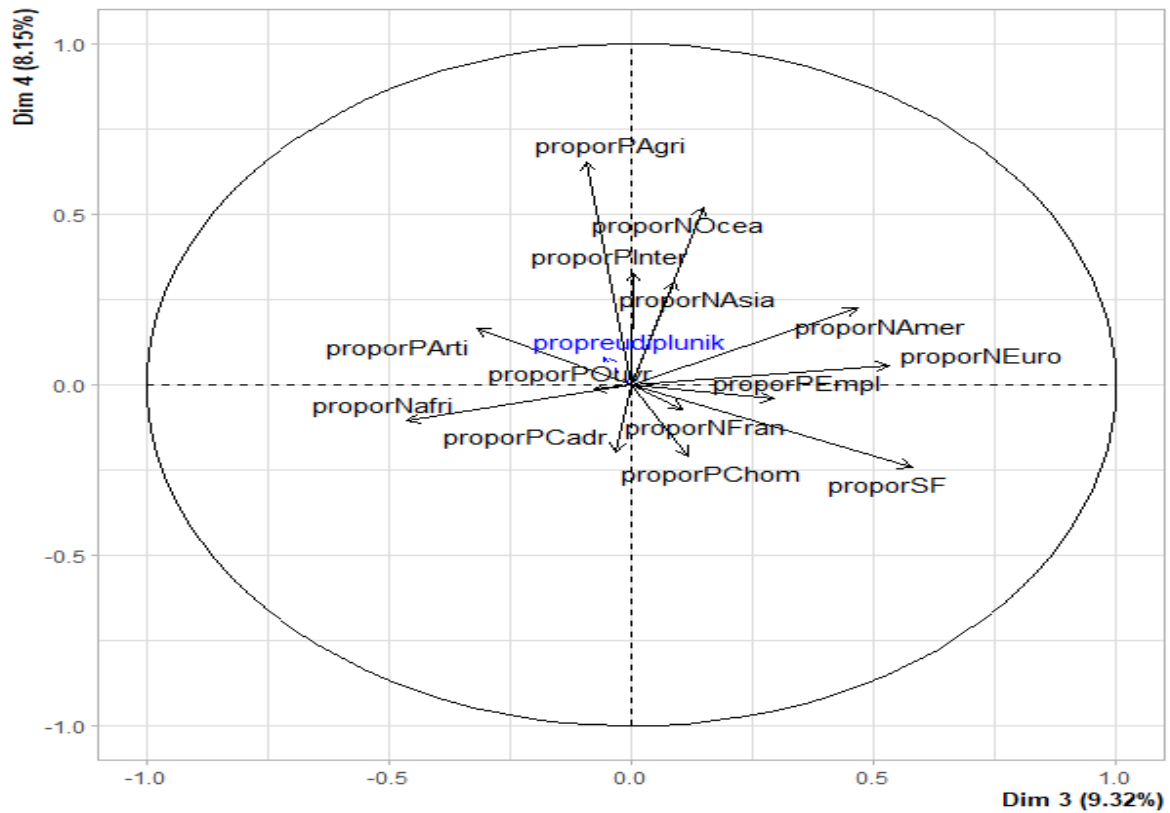
Graphique A.9 (2/2) Représentation factorielle des individus – Échantillon des formations uniques avec information complète



Graphique A.10 (1/2) Cercle des corrélations. Échantillon des formations uniques avec information complète



Graphique A.10 (2/2) Cercle des corrélations. Échantillon des formations uniques avec information complète



Graphique A.11 Dendrogramme de la classification ascendante hiérarchique Échantillon des formations uniques avec information complète

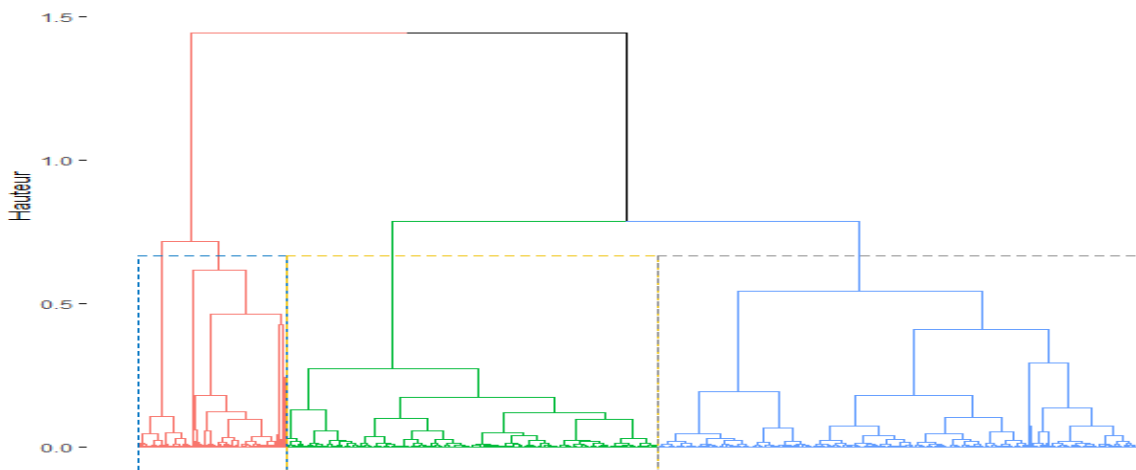


Tableau A.12 Récapitulatif de la classification ascendante hiérarchique. Échantillon des formations uniques avec information complète

Différence significative : * : 10%, ** : 5%, *** : 1%

Proportions	Total	Classe 1	Classe 2	Classe 3
Sexe				
Femmes	51.84	52.81	51.23	50.22
Nationalité				
Africains	6.75	4.20	4.47	37.94*
Américains	0.73	0.62	0.38	3.84
Asiatiques	1.65	1.44	0.84	8.47
Européens	1.68	1.58	1.06	6.53
Français	89.19	92.16	93.25	43.16**
Océaniens	0.01	0.00	0.01	0.05
CSP				
Agriculteurs	2.05	1.91	2.26	1.51
Artisans	8.78	10.01	7.46	10.39
Cadres	38.67	51.12**	28.75*	31.26
Chômeurs	7.48	4.58	8.19	20.23
Employés	14.64	11.41	18.01	11.10
Intermédiaires	12.98	11.10	15.48	7.19
Ouvriers	15.39	9.87	19.84	18.32
Taux de réussite	77.65	82.00	76.32	60.26
N(1)	152594	66485	75187	10922
N(2)	3970	1656	1933	381
Part redressée	40.29	46.86	38.22	14.60

**Tableau A.13 Régressions logistiques pour l'imputation des valeurs manquantes.
Variables sexe et nationalité (1/2)**

Les écarts types sont présentés entre parenthèse. Significativité : * : 10%, ** : 5%, *** : 1%

Effet marginal sur la probabilité d'être un étudiant	Homme	Étranger
Bac (réf = Economique)		
Littéraire	-0.134*** (0.001)	0.017*** (0.001)
Scientifique	0.072 *** (0.001)	0.002*** (0.000)
Technologique : STT	0.024*** (0.002)	0.016*** (0.001)
Technologique : Autre	0.106*** (0.002)	-0.005*** (0.001)
Professionnel	0.075*** (0.002)	0.042*** (0.001)
Dispense	0.027*** (0.001)	0.836*** (0.001)
Non renseigné	0.036*** (0.0125)	N.E
Discipline (réf = Sciences Humaines et Sociales)		
AES	0.072*** (0.003)	0.065*** (0.002)
Droit Sciences politiques	0.019*** (0.001)	0.021*** (0.001)
Langues	-0.041*** (0.002)	0.048*** (0.001)
Lettres et arts	0.001 (0.002)	0.017*** (0.001)
Médecine	-0.043*** (0.002)	0.001 (0.001)
Odontologie	0.086*** (0.006)	0.007** (0.003)
Pharmacie	0.005 (0.004)	0.018*** (0.002)
Pluri. droit éco	0.070*** (0.011)	0.102*** (0.007)
Pluri. lettres langues	-0.018*** (0.005)	0.014*** (0.002)
Pluri. santé	-0.096*** (0.002)	0.025*** (0.001)
Pluri. sciences	0.180*** (0.003)	0.033*** (0.001)
Sciences éco et gestion	0.134*** (0.001)	0.044*** (0.001)
Sciences appliquées	0.365*** (0.001)	0.042*** (0.001)
Sciences naturelles	0.003* (0.002)	0.018*** (0.001)
STAPS	0.333*** (0.002)	-0.018*** (0.001)

**Tableau A.13 Régressions logistiques pour l'imputation des valeurs manquantes.
Variables sexe et nationalité (2/2)**

Les écarts types sont présentés entre parenthèse. Significativité : * : 10%, ** : 5%, *** : 1%

Effet marginal sur la probabilité d'être un étudiant	Homme	Étranger
Niveau de la formation (réf = Bac +1)		
Bac+2	-0.022*** (0.001)	0.017*** (0.001)
Bac+3	-0.030*** (0.001)	0.019*** (0.001)
Bac+4	-0.028*** (0.001)	0.041*** (0.001)
Bac+5	-0.031*** (0.002)	0.049*** (0.001)
Formation(s) suivie(s) l'année précédente		
Même établissement	0.010** (0.005)	-0.002 (0.002)
Autre établissement	-0.007 (0.005)	-0.001 (0.002)
Même niveau de formation	0.048*** (0.005)	0.006*** (0.002)
Un an de moins	0.014*** (0.005)	-0.022*** (0.002)
Au moins un an de plus	0.082*** (0.005)	0.015*** (0.003)
Plus d'un an de moins	0.022*** (0.005)	-0.006*** (0.002)
Même discipline	-0.011** (0.004)	0.004* (0.002)
Autre discipline	-0.030*** (0.004)	-0.002 (0.002)
Réussite : oui	-0.057*** (0.008)	-0.015*** (0.000)
Age	0.002*** (0.000)	-0.003*** (0.000)
Nombre de formations suivies	-0.009*** (0.001)	-0.005*** (0.001)
Pseudo R2	0.099	0.53
Nombre d'observations	1567788	1566290

Tableau A.14 Régressions logistiques multinomiales pour l'imputation des valeurs manquantes. Variable nationalité étrangère (1/2)

Les écarts types sont présentés entre parenthèse. Significativité : * : 10%, ** : 5%, *** : 1%

Effet marginal sur la probabilité d'être un étudiant	Nationalité (réf = Afrique)			
	Amérique	Asie	Europe	Océanie
Bac (réf = Economique)				
Littéraire	-0.005 (0.004)	-0.000 (0.004)	0.058*** (0.006)	0.000 (0.000)
Scientifique	-0.024*** (0.003)	0.048*** (0.003)	0.010** (0.005)	0.001*** (0.000)
Technologique : STT	-0.030*** (0.004)	0.004 (0.005)	-0.050*** (0.007)	0.001** (0.001)
Technologique : Autre	-0.003 (0.006)	0.038*** (0.006)	0.041*** (0.009)	0.001 (0.001)
Professionnel	-0.105** (0.005)	0.023*** (0.005)	-0.078*** (0.007)	0.000 (0.000)
Dispense	-0.009*** (0.003)	0.125*** (0.003)	-0.086*** (0.004)	0.002*** (0.000)
Discipline (réf = Sciences Humaines et Sociales)				
AES	-0.019*** (0.004)	-0.024*** (0.005)	-0.182*** (0.005)	-0.003*** (0.000)
Droit Sciences politiques	-0.006** (0.002)	-0.030*** (0.003)	0.020*** (0.004)	-0.001** (0.000)
Langues	0.013*** (0.003)	0.065*** (0.004)	0.067*** (0.004)	0.000 (0.001)
Lettres et arts	0.055*** (0.003)	0.117*** (0.004)	0.011** (0.005)	0.001* (0.001)
Médecine	-0.062*** (0.003)	-0.048*** (0.005)	0.080*** (0.008)	-0.002*** (0.001)
Odontologie	-0.060*** (0.012)	0.030 (0.025)	0.091*** (0.030)	-0.003*** (0.000)
Pharmacie	-0.063*** (0.006)	0.023* (0.012)	-0.108*** (0.012)	-0.002 (0.001)
Pluri. droit éco	-0.034*** (0.011)	0.062*** (0.019)	-0.123*** (0.016)	-0.003*** (0.000)
Pluri. lettres langues	0.023*** (0.008)	-0.001 (0.009)	0.233*** (0.012)	-0.003*** (0.000)
Pluri. santé	-0.039*** (0.005)	-0.007 (0.006)	-0.078*** (0.007)	-0.003*** (0.000)
Pluri. sciences	-0.021*** (0.005)	0.010 (0.006)	-0.117*** (0.007)	-0.001 (0.001)
Sciences éco et gestion	-0.043*** (0.002)	0.053*** (0.003)	-0.126*** (0.003)	-0.001 (0.000)
Sciences appliquées	-0.048*** (0.002)	0.007** (0.003)	-0.192*** (0.003)	-0.003*** (0.000)
Sciences naturelles	-0.019*** (0.003)	-0.014*** (0.004)	-0.111*** (0.005)	-0.000 (0.000)
STAPS	-0.015** (0.007)	-0.081*** (0.007)	0.041*** (0.012)	-0.003*** (0.001)

Tableau A.14 Régressions logistiques multinomiales pour l'imputation des valeurs manquantes. Variable nationalité étrangère (2/2)

Les écarts types sont présentés entre parenthèse. Significativité : * : 10%, ** : 5%, *** : 1%

Effet marginal sur la probabilité d'être un étudiant	Nationalité (réf = Afrique)			
	Amérique	Asie	Europe	Océanie
Niveau de la formation (réf = Bac +1)				
Bac+2	-0.010*** (0.002)	-0.074*** (0.003)	-0.002 (0.003)	-0.001** (0.000)
Bac+3	-0.008*** (0.002)	-0.063*** (0.003)	0.003 (0.003)	-0.000 (0.000)
Bac+4	0.006*** (0.002)	-0.082*** (0.003)	-0.019*** (0.003)	-0.001*** (0.000)
Bac+5	0.008*** (0.002)	-0.069*** (0.003)	-0.017*** (0.003)	-0.002*** (0.000)
Formation(s) suivie(s) l'année précédente				
Même établissement	0.10 (0.009)	0.025** (0.011)	-0.086*** (0.013)	0.002 (0.003)
Autre établissement	0.001 (0.009)	0.009 (0.012)	-0.113*** (0.009)	-0.000 (0.002)
Même niveau de formation	-0.024*** (0.007)	-0.041*** (0.010)	-0.014 (0.012)	-0.001 (0.001)
Un an de moins	-0.016** (0.008)	-0.004 (0.011)	0.076*** (0.013)	0.000 (0.002)
Au moins un an de plus	-0.033*** (0.006)	-0.034*** (0.010)	-0.039*** (0.012)	-0.002*** (0.001)
Plus d'un an de moins	0.024** (0.011)	0.149*** (0.016)	0.008 (0.014)	0.000 (0.002)
Même discipline	-0.013 (0.009)	0.015 (0.011)	0.016 (0.012)	-0.002 (0.002)
Autre discipline	-0.005 (0.008)	0.037*** (0.012)	-0.005 (0.012)	-0.000 (0.002)
Réussite : oui	0.008*** (0.001)	0.042*** (0.002)	0.015*** (0.002)	-0.000 (0.000)
Age	0.002*** (0.002)	-0.002*** (0.000)	-0.002*** (0.000)	-0.000 (0.000)
Nombre de formations suivies	0.009*** (0.002)	0.017*** (0.003)	0.029*** (0.003)	0.000 (0.000)
Pseudo R2	0.073			
Nombre d'observations	215306			

Tableau A.15 Régressions logistiques multinomiales pour l'imputation des valeurs manquantes. Variable catégorie socioprofessionnelle (1/2)

Les écarts types sont présentés entre parenthèse. Significativité : * : 10%, ** : 5%, *** : 1%

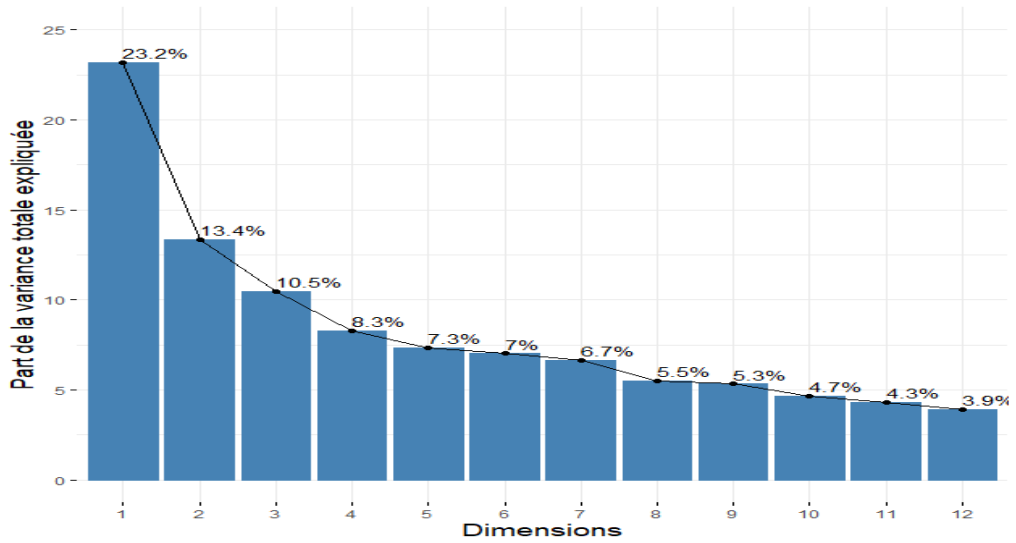
Effet marginal sur la probabilité d'être un étudiant	Catégorie socioprofessionnelle (réf = Cadre)					
	Agri.	Artisan	Employé	Sans prof.	Ouvrier	Prof. inter
Bac (réf = Economique)						
Littéraire	-0.004*** (0.000)	-0.006*** (0.001)	0.014*** (0.001)	0.023*** (0.001)	0.003*** (0.001)	-0.003*** (0.001)
Scientifique	-0.000 (0.000)	-0.006*** (0.001)	-0.024*** (0.001)	-0.005*** (0.001)	-0.028*** (0.001)	-0.006*** (0.001)
Technologique : STT	-0.002*** (0.000)	0.003** (0.001)	0.030*** (0.002)	0.054*** (0.001)	0.080*** (0.002)	-0.010*** (0.001)
Technologique : Autre	0.003*** (0.001)	-0.004*** (0.001)	0.018*** (0.002)	0.028*** (0.001)	0.051*** (0.002)	0.016*** (0.001)
Professionnel	0.002*** (0.001)	-0.006*** (0.001)	0.025*** (0.002)	0.115*** (0.002)	0.113*** (0.002)	-0.022*** (0.002)
Dispense	-0.001** (0.000)	0.039*** (0.001)	-0.024*** (0.001)	0.0053*** (0.001)	-0.019*** (0.001)	-0.054*** (0.001)
Discipline (réf = Sciences Humaines et Sociales)						
AES	-0.001 (0.001)	0.011*** (0.002)	-0.003 (0.002)	0.049*** (0.002)	0.035*** (0.002)	-0.020*** (0.002)
Droit Sciences politiques	-0.004*** (0.000)	0.018*** (0.001)	-0.025*** (0.001)	0.003*** (0.001)	-0.032*** (0.001)	-0.026*** (0.001)
Langues	-0.004*** (0.000)	0.004*** (0.001)	0.011*** (0.001)	0.018*** (0.001)	0.022*** (0.001)	-0.011*** (0.001)
Lettres et arts	-0.005*** (0.000)	-0.001 (0.001)	-0.013*** (0.001)	-0.002** (0.001)	-0.025*** (0.001)	-0.004*** (0.001)
Médecine	0.001 (0.001)	0.011*** (0.001)	-0.039*** (0.002)	-0.023*** (0.001)	-0.046*** (0.002)	-0.011*** (0.002)
Odontologie	-0.004** (0.002)	0.030*** (0.004)	-0.051*** (0.005)	-0.018*** (0.004)	-0.062*** (0.004)	-0.034*** (0.004)
Pharmacie	0.003** (0.001)	0.026*** (0.003)	-0.039*** (0.003)	-0.019*** (0.002)	-0.026*** (0.003)	-0.021*** (0.003)
Pluri. droit éco	-0.015*** (0.001)	0.013** (0.007)	-0.009 (0.008)	0.071*** (0.008)	0.029*** (0.009)	-0.029*** (0.007)
Pluri. lettres langues	-0.001 (0.001)	-0.010*** (0.003)	-0.041*** (0.003)	0.012*** (0.003)	-0.033*** (0.003)	-0.017*** (0.003)
Pluri. santé	0.000 (0.001)	0.015*** (0.001)	-0.035*** (0.002)	-0.013*** (0.001)	-0.025*** (0.002)	-0.015*** (0.002)
Pluri. sciences	-0.004*** (0.001)	0.004*** (0.002)	-0.029*** (0.002)	-0.011*** (0.001)	-0.033*** (0.02)	-0.033*** (0.002)
Sciences éco et gestion	0.001* (0.000)	0.034*** (0.010)	-0.023*** (0.001)	0.002** (0.008)	-0.008*** (0.001)	-0.022*** (0.001)
Sciences appliquées	-0.001** (0.000)	0.010*** (0.001)	-0.023*** (0.001)	-0.010*** (0.001)	0.001 (0.001)	-0.002** (0.001)
Sciences naturelles	0.010*** (0.001)	0.006*** (0.001)	-0.003 (0.002)	-0.003*** (0.001)	0.010*** (0.002)	0.001 (0.001)
STAPS	-0.000 (0.001)	0.004*** (0.001)	0.011*** (0.002)	-0.020*** (0.001)	0.006*** (0.002)	0.015*** (0.002)

Tableau A.15 Régressions logistiques multinomiales pour l'imputation des valeurs manquantes. Variable catégorie socioprofessionnelle (2/2)

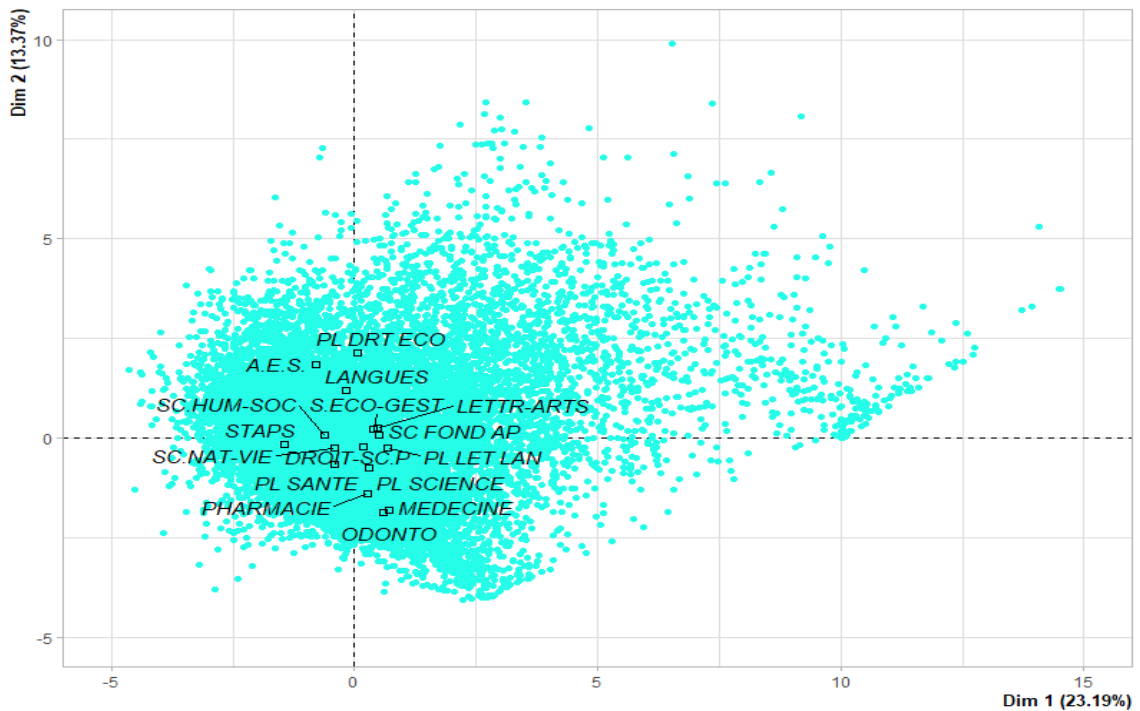
Les écarts types sont présentés entre parenthèse. Significativité : * : 10%, ** : 5%, *** : 1%

Effet marginal sur la probabilité d'être un étudiant	Catégorie socioprofessionnelle (réf = Cadre)					
	Agri.	Artisan	Employé	Sans prof.	Ouvrier	Prof. inter
Niveau de la formation (réf = Bac +1)						
Bac+2	0.004*** (0.000)	-0.002 (0.001)	-0.015*** (0.001)	-0.013*** (0.001)	-0.009*** (0.001)	-0.003*** (0.001)
Bac+3	0.006*** (0.000)	0.002* (0.001)	-0.016*** (0.001)	-0.018*** (0.001)	-0.006*** (0.001)	-0.003*** (0.001)
Bac+4	0.005*** (0.000)	-0.002 (0.001)	-0.019*** (0.001)	-0.022*** (0.001)	-0.010*** (0.001)	-0.005*** (0.001)
Bac+5	0.006*** (0.000)	-0.001 (0.001)	-0.023*** (0.001)	-0.030*** (0.001)	-0.015*** (0.001)	-0.006 (0.001)***
Formation(s) suivie(s) l'année précédente						
Même établissement	-0.002 (0.001)	-0.002 (0.003)	0.007* (0.004)	0.006** (0.003)	0.018*** (0.004)	0.004 (0.004)
Autre établissement	-0.002 (0.001)	-0.001 (0.003)	0.005 (0.004)	0.010*** (0.003)	0.006 (0.004)	-0.003 (0.004)
Même niveau de formation	-0.001 (0.001)	-0.005 (0.003)	0.011** (0.004)	0.008** (0.003)	0.021*** (0.004)	0.001 (0.004)
Un an de moins	0.000 (0.001)	-0.005* (0.003)	0.011*** (0.004)	-0.004 (0.003)	0.015*** (0.004)	0.006* (0.004)
Au moins un an de plus	0.002 (0.002)	-0.006* (0.003)	0.000 (0.004)	-0.138*** (0.003)	0.000 (0.004)	0.004 (0.004)
Plus d'un an de moins	-0.000 (0.002)	-0.001 (0.003)	0.002 (0.005)	-0.004 (0.003)	-0.005 (0.004)	-0.000 (0.004)
Même discipline	-0.001 (0.001)	0.008*** (0.003)	-0.010*** (0.003)	-0.002 (0.003)	-0.022*** (0.004)	-0.005 (0.003)
Autre discipline	-0.001 (0.001)	0.0004 (0.003)	-0.013*** (0.003)	-0.002 (0.003)	-0.022*** (0.003)	-0.007** (0.003)
Réussite : oui	0.004*** (0.000)	-0.001*** (0.001)	-0.005 (0.001)	-0.025*** (0.005)	-0.010*** (0.001)	0.007*** (0.001)
Age	0.000*** (0.000)	0.001*** (0.000)	-0.006*** (0.000)	0.003*** (0.000)	0.004*** (0.000)	-0.003*** (0.000)
Nombre de formations suivies	-0.002*** (0.000)	0.003*** (0.008)	-0.023*** (0.001)	-0.010*** (0.001)	-0.034*** (0.001)	-0.006*** (0.001)
Pseudo R2	0.026					
Nombre d'observations	1423740					

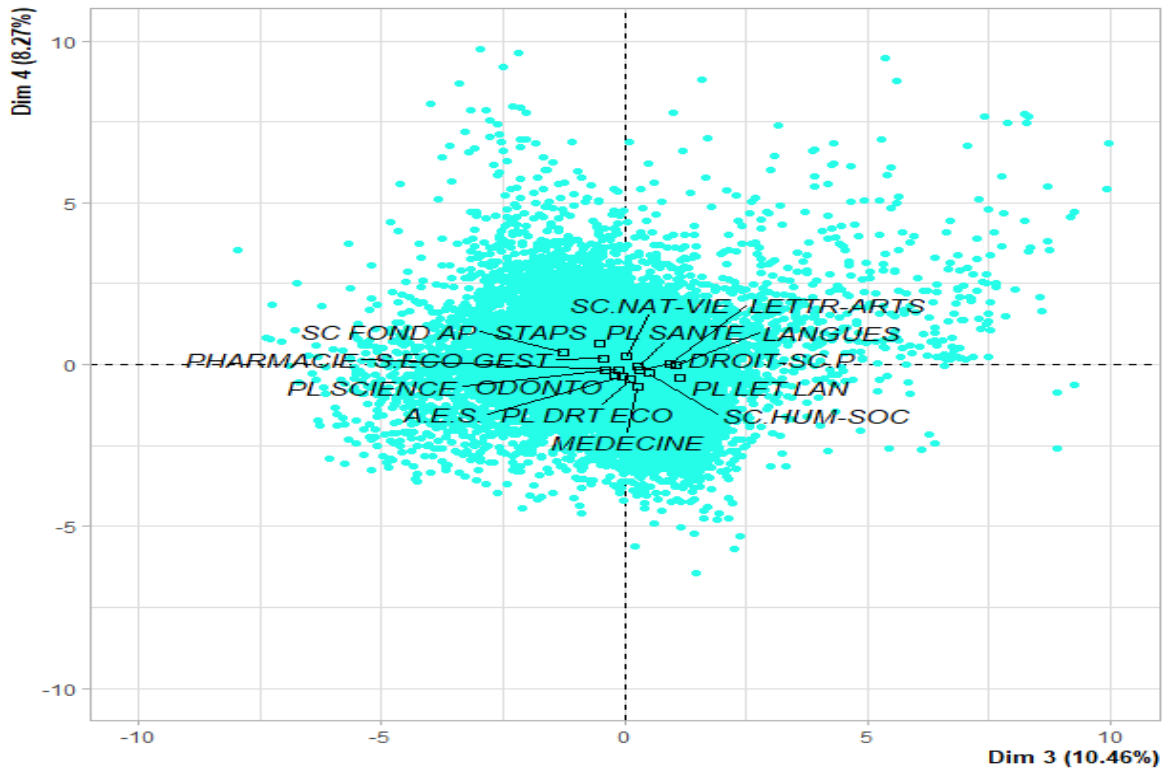
Graphique A.16 Distribution des valeurs propres – Échantillon des formations uniques avec informations manquantes imputées



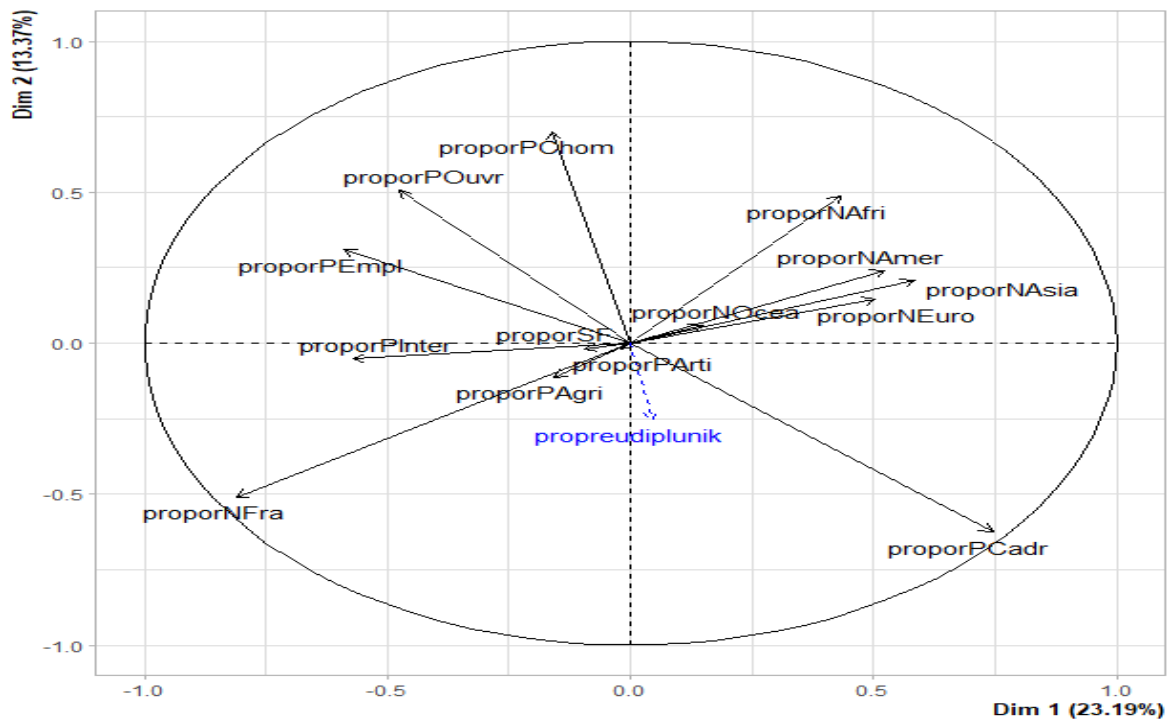
Graphique A.17 (1/2) Représentation factorielle des individus – Échantillon des formations uniques avec informations manquantes imputées



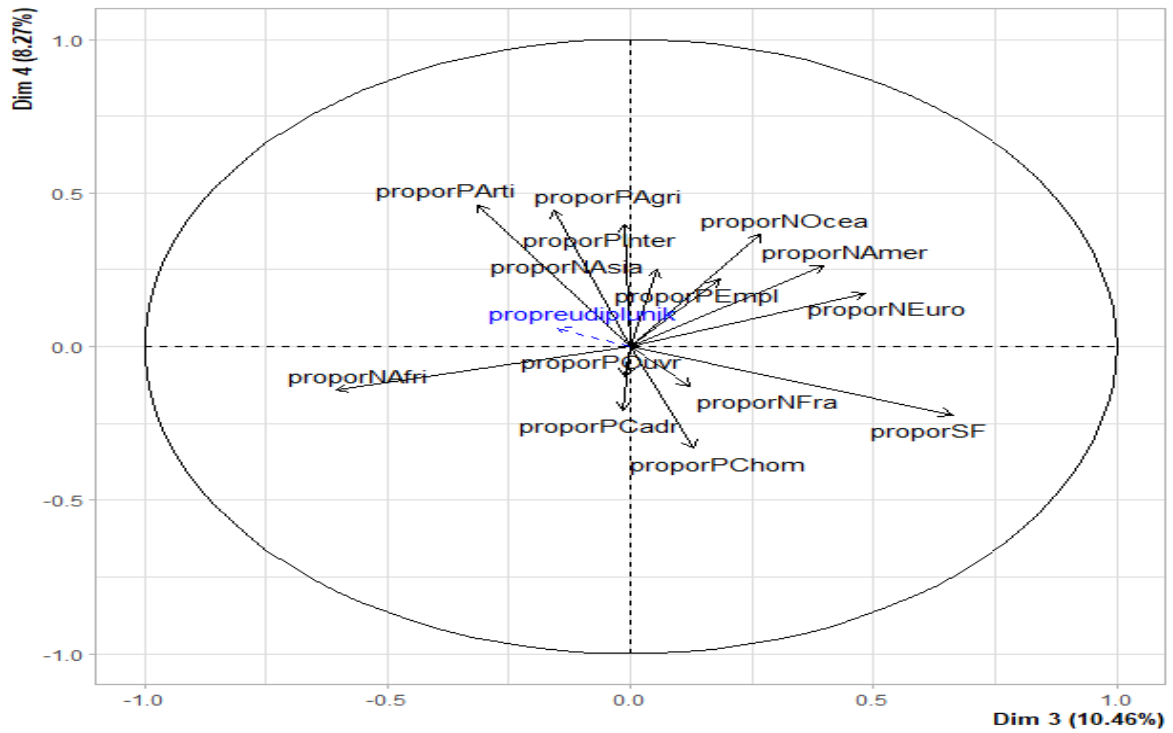
Graphique A.17 (2/2) Représentation factorielle des individus – Échantillon des formations uniques avec informations manquantes imputées



Graphique A.18 (1/2) Cercle des corrélations. Échantillon des formations uniques avec informations manquantes imputées



Graphique A.18 (2/2) Cercle des corrélations. Échantillon des formations uniques avec informations manquantes imputées



Graphique A.19 Dendrogramme de la classification ascendante hiérarchique Échantillon des formations uniques avec informations manquantes imputées

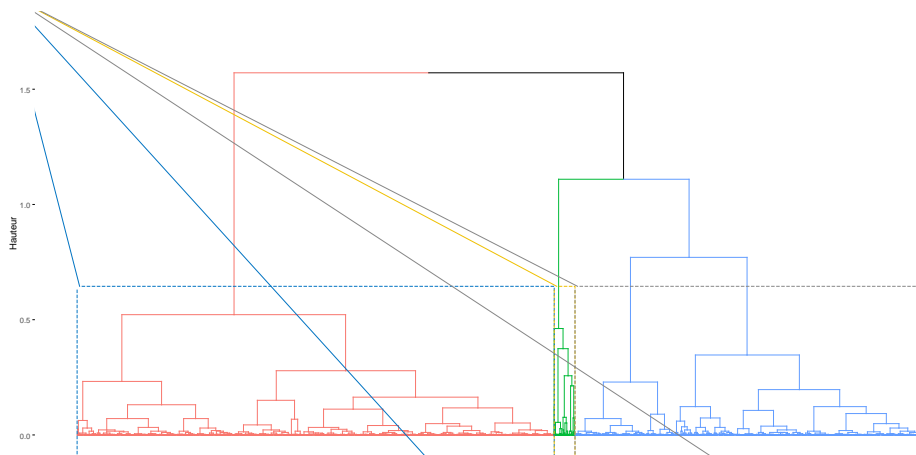


Tableau A.20 Récapitulatif de la classification ascendante hiérarchique. Échantillon des formations uniques avec informations manquantes imputées

Différence significative : * : 10%, ** : 5%, *** : 1%

Proportions	Total	Cluster 1	Cluster 2	Cluster 3
Sexe				
Femmes	56.77	58.74	50.17	65.84
Nationalité				
Africains	7.11	4.04***	16.67***	4.81
Américains	1.06	0.55**	1.75	15.13***
Asiatiques	2.45	1.17***	4.56*	31.76***
Européens	2.97	2.04**	4.07	30.98***
Français	86.37	92.19***	72.91***	16.29***
Océaniens	0.03	0.01	0.02	1.01**
CSP				
Agriculteurs	1.63	1.80	1.14	0.84
Artisans	8.44	8.42	8.66	5.93
Cadres	42.83	37.54***	57.23***	72.00**
Chômeurs	7.27	7.60	6.35	5.56
Employés	14.05	15.87***	8.96***	6.48
Intermédiaires	10.78	12.30***	6.53***	3.81
Ouvriers	15.00	16.47***	11.12***	5.37
Taux de réussite	65.54	64.20	70.70	49.02
N(1)	1576174	1170634	381577	23963
N(2)	18361	12266	5556	539
Part redressée	44.59	50.34	27.42	37.19

Tableau A.21 Régressions logistiques pour l'analyse de la réussite étudiante. Analyse globale (1/2)

Les écarts types sont présentés entre parenthèse. Significativité : * : 10%, ** : 5%, *** : 1%

Effet marginal sur la réussite étudiante	Spécification			
	(1)	(2)	(3)	(4)
Sexe (réf = Femme)				
Homme	-0.039*** (0.003)	-0.032*** (0.002)	-0.051*** (0.002)	-0.053*** (0.001)
Non renseigné	0.052 (0.059)	-0.049 (0.072)	-0.100 (0.077)	-0.094 (0.072)
Nationalité (réf = Français)				
Afrique	-0.081*** (0.004)	-0.053*** (0.004)	-0.065*** (0.004)	-0.066*** (0.004)
Amérique	-0.074*** (0.013)	-0.023** (0.011)	-0.013 (0.010)	-0.018* (0.009)
Asie	-0.008 (0.009)	0.026*** (0.008)	0.018** (0.007)	0.013* (0.007)
Europe	-0.071*** (0.009)	-0.029*** (0.007)	-0.022*** (0.007)	-0.028*** (0.006)
Océanie	-0.104*** (0.038)	-0.052* (0.030)	-0.043 (0.028)	-0.020 (0.028)
Non renseigné	0.246*** (0.023)	0.070 (0.046)	0.077* (0.045)	0.083 (0.043)
CSP (réf = Cadre)				
Agriculteur	0.021*** (0.004)	0.018*** (0.004)	0.014*** (0.003)	0.013*** (0.003)
Artisan	-0.034*** (0.002)	-0.018*** (0.002)	-0.021*** (0.002)	-0.020*** (0.002)
Employé	-0.057*** (0.003)	-0.023*** (0.002)	-0.023*** (0.02)	-0.022*** (0.002)
Sans profession	-0.156*** (0.005)	-0.089*** (0.003)	-0.086*** (0.003)	-0.076*** (0.003)
Ouvrier	-0.065*** (0.003)	-0.036*** (0.002)	-0.036*** (0.002)	-0.033*** (0.002)
Profession intermédiaire	-0.022*** (0.002)	-0.005** (0.002)	-0.006*** (0.002)	-0.006*** (0.002)
Non renseigné	-0.100*** (0.007)	-0.063*** (0.006)	-0.076*** (0.005)	-0.077*** (0.005)
Bac (réf = Economique)				
Littéraire	-	-0.057*** (0.003)	-0.031*** (0.002)	-0.031*** (0.002)
Scientifique	-	0.000 (0.005)	0.007*** (0.003)	0.011*** (0.002)
Technologique : STT	-	-0.099*** (0.004)	-0.110 (0.003)	-0.107*** (0.004)
Technologique : Autre	-	-0.063 (0.004)	-0.070*** (0.003)	-0.070*** (0.003)
Professionnel	-	-0.192*** (0.005)	-0.189*** (0.005)	-0.188*** (0.005)
Dispense	-	-0.067*** (0.005)	-0.060*** (0.005)	-0.066*** (0.004)
Non renseigné		0.194*** (0.037)	0.203*** (0.039)	0.189*** (0.038)

Tableau A.21 Régressions logistiques pour l'analyse de la réussite étudiante. Analyse globale (2/2)

Les écarts types sont présentés entre parenthèse. Significativité : * : 10%, ** : 5%, *** : 1%

Effet marginal sur la réussite étudiante	Spécification			
	(1)	(2)	(3)	(4)
Niveau de la formation (réf = Bac +1)				
Bac+2	-	0.045*** (0.011)	-0.014 (0.011)	-0.011 (0.010)
Bac+3	-	0.192*** (0.012)	0.121*** (0.012)	0.121*** (0.012)
Bac+4	-	0.093*** (0.013)	0.027** (0.012)	0.045*** (0.012)
Bac+5	-	0.156*** (0.015)	0.079*** (0.015)	0.103*** (0.014)
Formation(s) suivie(s) l'année précédente				
Même établissement	-	-0.015 (0.013)	-0.002 (0.013)	-0.009 (0.013)
Autre établissement	-	-0.015 (0.013)	0.002 (0.012)	0.008 (0.012)
Même niveau de formation	-	0.021 (0.013)	0.022* (0.013)	0.030** (0.013)
Un an de moins	-	0.149*** (0.014)	0.137*** (0.013)	0.137*** (0.014)
Au moins un an de plus	-	0.042*** (0.015)	0.025* (0.014)	0.024* (0.015)
Plus d'un an de moins	-	-0.104*** (0.020)	-0.088*** (0.019)	-0.095*** (0.020)
Même discipline	-	0.001 (0.013)	0.001 (0.012)	-0.003 (0.012)
Autre discipline	-	0.017 (0.012)	-0.001 (0.011)	-0.004 (0.012)
Age vis-à-vis de la formation				
Plus jeune que l'âge modal	-	-0.014*** (0.004)	-0.021*** (0.004)	-0.018*** (0.003)
Plus vieux que l'âge modal	-	-0.062*** (0.003)	-0.056*** (0.003)	-0.052*** (0.002)
Résultat redressé : oui	-	-0.057*** (0.010)	-0.082*** (0.011)	-0.088*** (0.011)
Nombre de formations suivies	-	-0.094*** (0.005)	-0.096*** (0.005)	-0.099*** (0.005)
ESPE : oui	-	-	-	-0.142*** (0.020)
Discipline (réf = Sciences Humaines et Sociales)	NON	NON	OUI	OUI
Etablissement (réf = Université de Lille)	NON	NON	NON	OUI
Pseudo R2	0.013	0.101	0.119	0.127
Nombre d'observations	1579015			

Tableau A.22 Distribution des effets mesurés dans les régressions par discipline, établissement et formation unique. Effets significatifs à 5%

Lecture : 77.78% des disciplines donnent lieu à l'estimation d'un effet marginal négatif et significatif à 5% d'être un homme. 18.52% à un effet non significatif et 3.70% à un effet positif et significatif à 5%

Critère/Niveau d'analyse	Disciplines				Etablissements				Formations uniques			
	N	-	0	+	N	-	0	+	N	-	0	+
Sexe : Homme	54	77.78	18.52	3.70	111	76.58	23.42	0.00	11331	14.16	82.26	3.58
Nationalité : Afrique	54	64.81	33.33	1.85	104	52.88	47.12	0.00	6037	17.57	72.88	9.54
Nationalité : Amérique	52	13.46	75.00	11.54	101	10.89	81.19	7.92	1153	9.89	79.10	11.01
Nationalité : Asie	53	3.77	84.91	11.32	90	8.89	81.11	10.00	2269	11.72	76.33	11.94
Nationalité : Europe	54	20.37	66.67	12.96	104	15.38	73.08	11.54	2773	8.94	79.34	11.72
Nationalité : Océanie	27	3.70	88.89	7.41	29	20.69	68.87	10.34	17	17.65	70.59	11.76
CSP : Agriculteur	53	3.77	84.91	11.32	110	1.82	83.64	14.55	1944	5.45	91.98	2.57
CSP : Artisan	54	25.93	74.07	0.00	111	18.02	81.08	0.90	6506	5.99	90.73	3.27
CSP : Employé	54	22.22	74.07	3.70	110	20.91	73.64	5.45	7595	6.04	90.35	3.61
CSP : Sans profession	54	72.22	27.78	0.00	111	66.67	31.53	1.80	5797	10.42	86.53	3.05
CSP : Ouvrier	54	44.44	55.56	0.00	110	33.64	64.55	1.82	8032	7.06	89.54	3.40
CSP : Intermédiaire	54	11.11	83.33	5.56	110	10.91	85.45	3.64	6956	5.09	91.63	3.28
N*	54				111				12885			
N(2)	1579015				1579015				1437567			

Tableau A.23 Distribution des effets mesurés dans les régressions par discipline, établissement et formation unique. Effets significatifs à 10%

Lecture : 83.33% des disciplines donnent lieu à l'estimation d'un effet marginal négatif et significatif à 10% d'être un homme. 12.96% à un effet non significatif et 3.70% à un effet positif et significatif à 10%

Critère/Niveau d'analyse	Disciplines				Etablissements				Formations uniques			
	N	-	0	+	N	-	0	+	N	-	0	+
Sexe : Homme	54	83.33	12.96	3.70	111	79.28	20.72	0.00	11331	19.34	75.39	5.28
Nationalité : Afrique	54	68.52	29.63	1.85	104	59.62	40.38	0.00	6037	22.25	66.56	11.20
Nationalité : Amérique	52	15.38	69.23	15.38	101	15.84	75.25	8.91	1153	12.49	73.63	13.88
Nationalité : Asie	53	11.32	73.58	15.09	90	11.11	76.67	12.22	2269	14.68	71.09	14.24
Nationalité : Europe	54	25.93	59.26	14.81	104	22.12	64.42	13.46	2773	11.32	74.43	14.24
Nationalité : Océanie	27	3.70	88.89	7.41	29	20.69	68.97	10.34	17	17.65	70.59	11.76
CSP : Agriculteur	53	3.77	83.02	13.21	110	4.55	77.27	18.18	1944	8.23	87.50	4.27
CSP : Artisan	54	31.48	68.52	0.00	111	24.32	73.87	1.80	6506	9.16	85.80	5.04
CSP : Employé	54	27.78	68.52	3.70	110	27.27	66.36	6.36	7595	9.36	85.25	5.39
CSP : Sans profession	54	81.48	18.52	0.00	111	67.57	29.57	2.70	5797	14.84	80.71	4.45
CSP : Ouvrier	54	53.70	46.30	0.00	110	40.00	57.27	2.73	8032	10.94	83.64	5.42
CSP : Intermédiaire	54	12.96	75.93	11.11	110	14.55	78.18	7.27	6956	8.18	86.21	5.61
N*	54				111				12885			
N(2)	1579015				1579015				1437567			

Tableau A.24 Méta-analyse des estimations par discipline, établissement et formation unique

Les écarts types sont présentés entre parenthèse. Significativité : * : 10%, ** : 5%, *** : 1%

Critère/Niveau d'analyse	Discipline	Etablissement	Formations uniques
Effet agrégé			
Sexe : Homme	-0.053*** (0.005)	-0.054*** (0.002)	-0.054*** (0.014)
Nationalité : Afrique	-0.077*** (0.009)	-0.073*** (0.007)	-0.112*** (0.004)
Nationalité : Amérique	-0.007 (0.015)	-0.014 (0.010)	-0.030*** (0.009)
Nationalité : Asie	0.008 (0.007)	0.003 (0.007)	-0.050*** (0.007)
Nationalité : Europe	-0.023* (0.012)	-0.018** (0.008)	-0.021*** (0.005)
Nationalité : Océanie	0.004 (0.028)	-0.079* (0.044)	-0.066 (0.085)
CSP : Agriculteur	0.012*** (0.004)	0.015*** (0.004)	-0.046*** (0.005)
CSP : Artisan	-0.016*** (0.003)	-0.015*** (0.002)	-0.030*** (0.002)
CSP : Employé	-0.014*** (0.003)	-0.015*** (0.002)	-0.023*** (0.002)
CSP : Sans profession	-0.055*** (0.004)	-0.059*** (0.003)	-0.069*** (0.002)
CSP : Ouvrier	-0.023*** (0.003)	-0.022*** (0.002)	-0.029*** (0.001)
CSP : Intermédiaire	-0.003 (0.003)	-0.004 (0.002)	-0.017*** (0.002)
N*	54	111	12885
N(2)	1579015	1579015	1437567

Tableau A.25 Méta-analyse des estimations par formations unique. Échantillon des formations uniques avec information concernant le résultat disponible

Le tableau se lit de la même façon que le tableau A.23 Les écarts types sont présentés entre parenthèse.

Significativité : * : 10%, ** : 5%, *** : 1%

Critère	N	-	0	+	Effet agrégé
Sexe : Homme	7934	16.99	77.28	5.73	-0.051*** (0.002)
Nationalité : Afrique	4244	24.84	64.59	10.58	-0.138*** (0.005)
Nationalité : Amérique	766	14.36	71.93	13.71	-0.055*** (0.012)
Nationalité : Asie	1502	16.78	70.17	13.05	-0.078*** (0.009)
Nationalité : Europe	1707	13.18	72.00	14.82	-0.046*** (0.007)
Nationalité : Océanie	7	14.29	57.14	28.57	-0.056 (0.158)
CSP : Agriculteur	1024	10.35	85.55	4.10	-0.071*** (0.007)
CSP : Artisan	3946	8.39	85.81	5.80	-0.027*** (0.003)
CSP : Employé	4710	8.30	86.35	5.35	-0.021*** (0.002)
CSP : Sans profession	3570	12.83	82.38	4.79	-0.064*** (0.003)
CSP : Ouvrier	5194	10.13	84.50	5.37	-0.027*** (0.002)
CSP : Intermédiaire	4163	7.85	86.43	5.72	-0.017*** (0.002)
N*			9139		
N(2)			768003		

Tableau A.26 Méta-régressions des estimations effectuées sur les formations uniques : effets du sexe et de la nationalité

Critères	Sexe : Homme		Nationalité : Afrique		Nationalité : Asie		Nationalité : Europe	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Contrôles								
Constante	-	-0.034** (0.015)	-	-0.029 (0.048)	-	-0.023 (0.081)	-	-0.106 (0.078)
Résultat redressé : oui	-0.009*** (0.003)	0.006 (0.004)	0.084*** (0.009)	0.041*** (0.014)	0.081*** (0.014)	0.038 (0.023)	0.057*** (0.011)	0.003 (0.020)
Ln(effectif)	-0.004** (0.001)	0.003* (0.002)	0.037*** (0.004)	0.045*** (0.005)	0.048*** (0.006)	0.054*** (0.008)	0.039*** (0.005)	0.047*** (0.007)
Taux de réussite	0.000*** (0.000)	0.000 (0.000)	-0.003*** (0.000)	-0.002*** (0.000)	-0.004*** (0.000)	-0.003*** (0.000)	-0.004*** (0.000)	-0.003*** (0.000)
% critère de référence	-0.001** (0.000)	-0.001*** (0.000)	-0.005*** (0.001)	-0.001 (0.001)	-0.003 (0.002)	-0.001 (0.002)	0.002 (0.002)	0.003 (0.002)
%critère de référence ²	-0.000 (0.000)	0.000 (0.000)	0.000*** (0.000)	-0.000 (0.000)	0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)	-0.000** (0.000)
% critère non renseigné	0.001 (0.000)	0.000 (0.005)	-0.002* (0.001)	-0.002 (0.001)	-0.004 (0.003)	-0.001 (0.003)	-0.005* (0.003)	0.002 (0.003)
Niveau de la formation (réf = Bac+1)								
Bac+2	0.015*** (0.004)	0.019*** (0.006)	-0.051*** (0.014)	0.022 (0.017)	-0.068*** (0.022)	0.039 (0.028)	-0.062*** (0.016)	0.010 (0.021)
Bac+3	0.015*** (0.004)	0.022*** (0.007)	-0.125*** (0.013)	-0.024 (0.019)	-0.121*** (0.022)	0.021 (0.033)	-0.085*** (0.016)	0.023 (0.027)
Bac+4	0.027*** (0.005)	0.037*** (0.006)	-0.103*** (0.013)	-0.019 (0.019)	-0.145*** (0.020)	-0.009 (0.032)	-0.087*** (0.016)	0.008 (0.026)
Bac+5	0.030*** (0.005)	0.039*** (0.007)	-0.085*** (0.013)	0.016 (0.021)	-0.098*** (0.020)	0.057 (0.035)	-0.103*** (0.016)	0.013 (0.028)
Etablissement : ESPE	-0.009 (0.009)	-0.011 (0.009)	-0.074** (0.040)	-0.064 (0.041)	-0.077 (0.075)	-0.082 (0.075)	-0.027 (0.030)	-0.025 (0.033)
Discipline (réf = SHS)								
Littéraire	-0.019*** (0.004)	-0.004 (0.005)	0.030** (0.013)	-0.008 (0.013)	0.090*** (0.018)	0.077*** (0.018)	0.065*** (0.013)	0.057*** (0.013)
Scientifique	0.019*** (0.003)	-0.002 (0.004)	0.043*** (0.009)	0.041*** (0.009)	0.012 (0.016)	0.024 (0.016)	0.004 (0.014)	-0.000 (0.014)
Médecine	0.017*** (0.006)	0.014** (0.007)	0.004 (0.022)	-0.017 (0.022)	-0.013 (0.036)	-0.104*** (0.036)	-0.107*** (0.026)	-0.143*** (0.027)
Pseudo R2	-	0.017	-	0.067	-	0.101	-	0.095

Tableau A.27 Méta-régressions des estimations effectuées sur les formations uniques : effets de la catégorie socioprofessionnelle

Critère	CSP : Agriculteur		CSP : Employé		CSP : Ouvrier		CSP : Sans profession	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Contrôles								
Constante	-	-0.025 (0.058)	-	-0.022 (0.018)	-	-0.001 (0.019)	-	-0.052* (0.027)
Résultat redressé : oui	0.047*** (0.010)	0.042** (0.019)	-0.006** (0.003)	0.003 (0.006)	-0.007** (0.003)	0.006 (0.006)	-0.014*** (0.005)	0.007 (0.009)
Ln(effectif)	0.040*** (0.005)	0.022*** (0.005)	0.006*** (0.002)	0.008*** (0.002)	0.003* (0.002)	0.004** (0.002)	0.007*** (0.003)	0.010*** (0.003)
Taux de réussite	-0.003*** (0.001)	-0.003*** (0.000)	-0.000* (0.000)	-0.000** (0.000)	-0.000 (0.000)	-0.000 (0.000)	-0.001*** (0.000)	-0.001 (0.000)
% critère de référence	-0.002 (0.002)	0.000 (0.002)	-0.002*** (0.001)	-0.002*** (0.001)	-0.003*** (0.001)	-0.003*** (0.001)	-0.002** (0.001)	-0.002** (0.001)
%critère de référence ²	0.000 (0.000)	0.000 (0.000)	0.000** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000* (0.000)	0.000** (0.000)
% critère non renseigné	-0.001 (0.000)	-0.000 (0.000)	-0.000** (0.000)	-0.001*** (0.000)	-0.000** (0.000)	-0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Niveau de la formation (réf = Bac+1)								
Bac+2	-0.049*** (0.014)	0.031 (0.020)	0.021*** (0.004)	0.030*** (0.006)	0.025*** (0.005)	0.036*** (0.007)	0.036*** (0.007)	0.063*** (0.010)
Bac+3	-0.055*** (0.014)	0.075*** (0.025)	0.013*** (0.005)	0.027*** (0.007)	0.018*** (0.005)	0.034*** (0.008)	0.028*** (0.007)	0.070*** (0.013)
Bac+4	-0.046*** (0.016)	0.037 (0.024)	0.010** (0.005)	0.023 (0.007)	0.004 (0.005)	0.018** (0.008)	0.013 (0.008)	0.046*** (0.012)
Bac+5	-0.085*** (0.016)	0.025 (0.028)	-0.002 (0.005)	0.014 (0.009)	-0.004 (0.006)	0.011 (0.009)	0.003 (0.008)	0.044*** (0.014)
Etablissement : ESPE	0.034* (0.020)	0.045* (0.025)	0.028*** (0.009)	0.023** (0.010)	0.030*** (0.009)	0.034*** (0.010)	0.020 (0.014)	0.012 (0.015)
Discipline (réf = SHS)								
Littéraire	-0.013 (0.016)	-0.033*** (0.017)	0.010** (0.005)	0.009* (0.005)	0.013** (0.005)	0.010** (0.005)	0.018*** (0.007)	0.013* (0.007)
Scientifique	-0.017 (0.012)	-0.012 (0.012)	-0.007* (0.004)	-0.003 (0.004)	0.001 (0.004)	0.006 (0.004)	0.000 (0.006)	0.004 (0.006)
Médecine	-0.035* (0.019)	-0.067*** (0.020)	-0.014** (0.017)	-0.015** (0.007)	-0.013* (0.008)	-0.005 (0.008)	-0.026** (0.013)	-0.029** (0.013)
Pseudo R2	-	0.110	-	0.015	-	0.032	-	0.018

Tableau A.28 : Méta-régressions des estimations effectuées sur les formations uniques : spécification avec la classe d'appartenance

Critères	Sexe : Homme		Nationalité : Afrique		Nationalité : Asie		Nationalité : Europe	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Contrôles								
Constante	-0.061*** (0.002)	-0.087*** (0.011)	-0.126*** (0.005)	-0.171*** (0.032)	-0.048*** (0.010)	-0.119** (0.058)	-0.017** (0.007)	-0.078*** (0.045)
Classe d'appartenance (réf = 1)								
Classe 2	0.023*** (0.003)	0.013*** (0.004)	0.031*** (0.008)	0.069*** (0.010)	-0.008 (0.014)	0.056*** (0.016)	-0.016 (0.012)	0.032** (0.013)
Classe 3	0.021* (0.013)	0.025* (0.013)	0.090*** (0.052)	0.126** (0.052)	0.079** (0.043)	0.174*** (0.043)	0.042 (0.035)	0.090** (0.036)
Résultat redressé : oui		0.007 (0.005)		0.033** (0.014)		0.028 (0.023)		-0.004 (0.019)
Ln(effectif)		0.001 (0.002)		0.037*** (0.005)		0.048*** (0.008)		0.045*** (0.006)
Taux de réussite		0.000 (0.000)		-0.002*** (0.000)		-0.004 (0.000)		-0.003*** (0.000)
Niveau de la formation (réf = Bac+1)								
Bac+2		0.016*** (0.006)		0.019 (0.017)		0.031 (0.028)		0.008 (0.021)
Bac+3		0.017** (0.007)		-0.024 (0.020)		0.016 (0.033)		0.024 (0.027)
Bac+4		0.028*** (0.006)		-0.013 (0.019)		-0.014 (0.033)		0.009 (0.026)
Bac+5		0.030*** (0.007)		0.017 (0.021)		0.048 (0.036)		0.013 (0.029)
Etablissement : ESPE		-0.013 (0.009)		-0.079** (0.040)		-0.093 (0.075)		-0.042 (0.033)
Discipline (réf = SHS)								
Littéraire		-0.016*** (0.004)		0.002 (0.013)		0.082*** (0.018)		0.063*** (0.013)
Scientifique		0.019*** (0.003)		0.045*** (0.009)		0.027* (0.016)		0.002 (0.014)
Médecine		0.012*** (0.007)		-0.035 (0.022)		-0.114*** (0.036)		-0.147*** (0.026)
Pseudo R2	0.011	0.025	0.003	0.060	0.006	0.100	0.009	0.092

Tableau A.29 Méta-régressions des estimations effectuées sur les formations uniques : spécification avec la classe d'appartenance

Critères	CSP : Agriculteur		CSP : Employé		CSP : Ouvrier		CSP : Sans profession	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Contrôles								
Constante	-0.039*** (0.005)	-0.022 (0.043)	-0.022*** (0.002)	-0.061*** (0.014)	-0.027*** (0.002)	-0.046*** (0.014)	-0.071*** (0.003)	-0.084*** (0.021)
Cluster d'appartenance (réf = 1)								
Cluster 2	-0.035*** (0.013)	-0.014 (0.015)	-0.003 (0.004)	-0.002 (0.005)	-0.012*** (0.004)	-0.007 (0.005)	0.008 (0.006)	0.010 (0.007)
Cluster 3	0.012 (0.073)	-0.035 (0.072)	-0.044** (0.021)	-0.045** (0.021)	-0.014 (0.022)	-0.014 (0.022)	-0.005 (0.036)	-0.017 (0.037)
Résultat redressé : oui		0.041** (0.018)		0.004 (0.005)		0.005 (0.006)		0.006 (0.009)
Ln(effectif)		0.022*** (0.006)		0.007*** (0.001)		0.002 (0.002)		0.008*** (0.003)
Taux de réussite		-0.003*** (0.000)		-0.000** (0.000)		-0.000* (0.000)		-0.001*** (0.000)
Niveau de la formation (réf = Bac+1)								
Bac+2		0.032* (0.019)		0.029** (0.006)		0.031*** (0.007)		0.059*** (0.010)
Bac+3		0.077*** (0.024)		0.026*** (0.007)		0.030*** (0.008)		0.066*** (0.013)
Bac+4		0.041* (0.023)		0.021*** (0.007)		0.012 (0.008)		0.038*** (0.012)
Bac+5		0.031 (0.027)		0.009 (0.009)		0.005 (0.009)		0.035** (0.014)
Etablissement : ESPE		0.038 (0.025)		0.024** (0.010)		0.038*** (0.010)		0.022 (0.015)
Discipline (réf = SHS)								
Littéraire		-0.033** (0.016)		0.011** (0.005)		0.012** (0.005)		0.014** (0.007)
Scientifique		-0.012 (0.012)		-0.004 (0.004)		0.003 (0.004)		0.001 (0.006)
Médecine		-0.069*** (0.020)		-0.016** (0.007)		-0.011 (0.008)		-0.031** (0.013)
Pseudo R2	0.006	0.112	0.005	0.018	-0.006	0.025	0.000	0.017