

ANALYSE DE L'EFFET SOCIO-DEMOGRAPHIQUE ET DU PARCOURS INITIAL SUR L'EFFICACITE INTERNE A L'UNIVERSITE MARIEN NGOUABI : LE CAS DE LA LICENCE 2 ET 3 A LA FACULTE DES SCIENCES ECONOMIQUES

Auguste M'PIAYI

Docteur en Sciences Economiques
Enseignant à la Faculté des Sciences Economiques
Université Marien Ngouabi,
Congo

melissampiayi@yahoo.fr

Résumé :

Cet article présente les résultats d'une recherche qui analyse l'impact sociodémographique des acquisitions initiales sur l'efficacité interne de l'enseignement supérieur congolais, à partir des données tirées d'une enquête auprès des étudiants de la Faculté des Sciences Economiques (FSE) à l'Université Marien NGOUABI (UMNG). Le modèle logistique confirme l'efficacité du baccalauréat série C, le baccalauréat technique étant le moins recommandé. La vie en couple, l'origine sociale aisée et la domiciliation hors résidence universitaire des étudiants auraient un effet bénéfique très significatif. Ces résultats recommandent une analyse approfondie des conditions de vie des campusards afin de mettre en œuvre une politique efficace de logement et d'aide sociale estudiantins. Ils plaident aussi en faveur d'une meilleure transition pédagogique lycée/université et de l'élargissement et de l'approfondissement des analyses d'évaluation du système.

Mots clés : enseignement universitaire, chances de réussite, réussite universitaire, efficacité, modèle logistique, Congo-Brazzaville.

Abstract:

This paper presents the results of a research that analyzes the socio-demographic impact and initial acquisitions on the internal efficiency of Congolese higher education, based on data drawn from a survey of students in the Faculty of Economics (FSE) at the Marien NGOUABI University. The logistical model confirms the effectiveness of the C baccalaureate, while the technical baccalaureate is the least recommended. Living with a partner, wealthier social origins and residence outside the university residence of students would have a very significant beneficial effect. These results recommend an in-depth analysis of the living conditions of the campus residents in order to implement an effective student housing and social assistance policy. They also plead in favor of a better high school / university pedagogical transition and finally the broadening and deepening of the evaluation analyzes of the system.

Keywords: university, education, chances of success, university success, efficiency, logistical model, Congo-Brazzaville.

Classification JEL Z0

1- Introduction

Les travaux sur l'évaluation de l'enseignement supérieur au Congo sont d'une façon générale relativement rares alors que de nombreux problèmes qui minent le système nécessitent des analyses régulières. C'est le cas de l'échec massif enregistré au sein de cette institution. L'enseignement universitaire public est celui qui alimente l'inefficacité de l'enseignement supérieur congolais (Menga-Mokombi, 2019).

On observe par ailleurs la massification de l'enseignement secondaire qui débouche directement sur les formations universitaires (notamment celles qui ne sont pas sélectives) du fait de la capacité des autres établissements (privés) encore insuffisamment importante et du coût des études qui freinent l'accueil d'un public plus large. Ce contexte assez particulier conduit à s'interroger sur la capacité des établissements publics à répondre à cette demande qui est sans cesse croissante. Le problème que nous posons ici est celui de l'efficacité ou plutôt de l'inefficacité interne de l'enseignement supérieur au Congo.

Pour éviter l'effet de saturation déjà visible en début du premier cycle qui, de fil en aiguille, contribue à la dégradation encore plus importante par le jeu de cercle vicieux et disloquerait l'Université au profit des établissements privés de l'enseignement supérieur, mettant de facto à l'écart des jeunes socialement défavorisés, il convient de déterminer les facteurs d'inefficacité interne des formations universitaires congolaises, en focalisant particulièrement l'attention sur les caractéristiques socio-démographiques et les acquisitions initiales déterminées à l'entrée de la formation universitaire. Il faut noter que l'échec universitaire n'est pas spécifique au cas du Congo. Il s'agit d'une situation relativement universelle mais qui s'est posée avec plus de sévérité dans des pays qui se sont engagés dans la voie de la démocratisation de l'éducation scolaire et universitaire. On peut citer le cas de la France où plus de la moitié des étudiants primo arrivants sont recalés à la fin de l'année. Ce travail déjà justifié par la nécessité de sauver l'université permettra de mettre en lumière les dysfonctionnements structurels, d'établir des indicateurs de pilotage qui aideraient à tracer des pistes de politique éducative.

Après la revue de la littérature centrée sur la notion d'efficacité interne il conviendra, par la suite, d'exposer successivement les données statistiques, la méthodologie, les résultats de l'analyse empirique et leur implication en matière de politique éducative.

2- Revue de la littérature

Sur le plan purement théorique, le débat peut porter sur la notion d'efficacité interne du système éducatif. Ce dernier permet de rendre compte de sa capacité à remplir ses fonctions qui, pour l'essentiel, concernent la transmission des connaissances, donc la production des hommes et de femmes formés. En ce sens, elle se confond avec celle de la productivité (M'piayi, 2001). Il s'agit en effet, d'établir un rapport entre les inputs de production et les outputs. Cette conception considérée par Kirschling (cité par De Ketele, 1989) et par les experts de la Banque mondiale (Psacharopoulos et Woodhall, 1988) intègre les dépenses réelles.

L'efficacité maximum est atteinte à partir du moment où il n'est plus possible d'améliorer le niveau de productivité. C'est la situation d'efficience (Paul, 2007). Mais il peut s'agir de mettre

face à face ce rapport de productivité et le niveau de réalisation des objectifs fixés a priori, comme le rappellent De Ketele et Sall (1997) en citant Legendre. L'efficacité interne se réduirait donc au chiffrage des résultats à la fin d'un cycle de formation ou encore au rapport entre les objectifs que se serait fixé le système éducatif et les résultats atteints. La question, au final, est donc celle de la mesure de ce résultat, c'est-à-dire de l'output. Il peut s'agir de l'output quantitatif, c'est le cas des abandons, des réussites ou passages dans un cycle supérieur, indicateurs utilisés souvent par les organisations internationales (Sall et al, 1997). L'output quantitatif, malgré son usage abondant, porte des limites dont celle de la qualité différenciée possible des entrées (les apprenants recrutés), entre établissements (Sall et al, 1997). Par exemple un établissement scolaire qui, à coût donné, produit davantage de diplômés qu'un autre qui assure la même formation, n'est pas nécessairement plus efficace ou efficient dès lors que les modalités de recrutement des élèves y sont plus sélectives.

Un établissement qui sélectionne à l'entrée les apprenants en fonction des prédispositions (cursus et niveau initial) produira de toute évidence plus de diplômés que l'établissement moins sélectif. Il serait donc scientifiquement erroné de déduire que les meilleurs taux de réussite observés au sein de l'établissement le plus sélectif serait le signe d'une plus grande efficacité productive.

La qualité des notions que les apprenants auraient réellement acquises à la sortie (objectifs pédagogiques) mesurée à partir des tests standardisés, rapportée à celle des notions dont ils seraient dotés à l'entrée (Bressoux, 1993), pourrait régler au moins en partie cette difficulté. Cette procédure mesure en réalité la valeur ajoutée par la formation, autrement dit le gain cognitif ou progression des apprenants entre le début et la fin de la formation.

Si, sur le plan théorique, ce raisonnement semble aisé, au niveau empirique, il est assez incongru de mesurer ce différentiel par la simple différence absolue, pour cause les compétences évaluées sur la période ne relèvent pas des mêmes thématiques. C'est la raison pour laquelle il est plus prompt selon Hanushek (1979) de déterminer plutôt l'impact du niveau initial sur la réussite finale à travers une fonction de production.

Au-delà de cette considération productiviste, l'efficacité notamment en éducation laisse entrevoir des pistes complémentaires d'évaluation. Un système éducatif efficace (au sens d'efficacité interne) peut aussi être celui qui réduit les inégalités dans les performances acquisitionnelles entre des catégories d'apprenants, par exemple entre les filles et les garçons, les étudiants d'origine sociale modeste et leurs collègues provenant des milieux plus aisés. Cette dimension sociale est évoquée par Bocquillon (2020) qui propose trois pistes d'idéologies pédagogiques d'interprétation de l'efficacité : l'égalité des chances, l'égalité de traitement et l'égalité des acquis. La première correspond à l'idée d'offrir à tous, les mêmes possibilités d'accéder à une formation de son choix sans effet discriminant. La seconde plaide pour une égalité dans les connaissances transmises ainsi que les processus de transmission eux-mêmes. Enfin, l'égalité des acquis évoque quant à elle la réduction de la variance entre les compétences acquises par les apprenants. Dans la même perspective, Deketele et Sall (1997) évoquent plutôt la notion d'équité de l'ordre éthique social, fractionnée en équité socio-économique, équité de confort (répartition non discriminatoire des étudiants dans les établissements n'appliquant pas les mêmes pédagogies et ne disposant pas des mêmes moyens), équité pédagogique (qui concerne les acquisitions), proposée aussi par Bressoux (1993) et équité de production (associée au nombre de diplômés produit au final). McMahon, repris par Psacharopoulos et Woodhall

(1988) distingue l'équité horizontale (même traitement aux individus égaux), équité verticale (traitement différencié aux individus inégaux et l'équité intergénérationnelle (de la reproduction sociale des inégalités).

En clair, un système éducatif efficace devrait gommer les discriminations en matière d'accès aux différents types de formation, aux compétences cognitives transmises et acquises.

L'efficacité peut aussi traduire la capacité du système éducatif à fonctionner sans heurts, sans conflits sociaux. En effet, les dysfonctionnements, le sentiment d'injustice... peuvent se traduire par des soulèvements sociaux qui abîmeraient davantage la qualité de la formation, donc son efficacité. La question de l'efficacité interne laisse ainsi une marge très importante à la fois dans sa définition et ses indicateurs de mesure.

Cette discussion montre combien la notion d'efficacité interne, quoi qu'importante, est multidimensionnelle et ne pourrait correspondre de façon consensuelle à une définition et une méthodologie au sein de la communauté scientifique. Cette difficulté est déjà annoncée Eicher, cité par Duguet et al (2016). Toutefois, compte tenu de leur apport complémentaire, des méthodes d'évaluation différentes peuvent être menées conjointement. La méthode des ratios (taux de réussite, taux d'abandon...) mesure l'efficacité globale du dispositif d'enseignement sans en démêler les effets individuels des facteurs.

Sur le plan de la revue empirique en matière d'évaluation des systèmes éducatifs, on peut déjà observer une attention très marquée des chercheurs sur l'effet des caractéristiques individuelles sur la réussite. Celle qui a été préalablement mise en avant du fait même de la démocratisation de l'enseignement dont le but consistait à conduire 80% d'une classe d'âge au niveau du baccalauréat. Ces analyses portent sur des indicateurs variés de la réussite universitaire. Qu'il s'agisse de la note, dite moyenne obtenue aux épreuves, des probabilités de valider une année (première, deuxième ou troisième année) sans retard ou non, de la capacité de rétention...

La réussite des premières années en droit, lettres et psychologie, observée à la faculté de Dijon en 1995, était plus favorable aux étudiants socialement favorisés (Duru-Bellat, 1995). Les travaux de Lemaire (2000) ont tendance à montrer que cette tonalité sociale baisse d'intensité en 2^{ème} année puisque l'écart des chances de passage en 3^{ème} année sans doubler, entre étudiants de père ouvrier ou employé et leurs autres collègues ne sont pas significatives. Mais l'effet spécifique de « père enseignant » est extrêmement important. Les travaux réalisés par Felouzi (2000) et Michaut (2000), sur les étudiants des sites de Dijon, Nantes et Toulouse confirment cette sensibilité assez variable, voire de moins en moins marquée de l'effet de l'origine sociale sur la réussite universitaire, du fait très certainement de la sévérité de la sélection pratiquée en amont notamment en fin de cycle lycéen et en début de cycle universitaire. Cette observation peut laisser croire au caractère progressif d'une modération des objectifs de la démocratisation de l'enseignement supérieur.

Les auteurs reconnaissent que l'effet de l'origine sociale s'observe sur le choix de poursuivre les études après avoir franchi un niveau donné sans pour autant conduire à la différenciation de la réussite, toutes choses égales par ailleurs. Goux et Maurin (1995) évoquent le phénomène de la démocratisation uniforme, pour montrer que l'accès massif des étudiants socialement défavorisés à l'enseignement supérieur, qui justifie les notions de démocratisation et de massification, masque la persistance des discriminations qui sont désormais transférées dans les

modes de recrutement au sein des établissements. Les différences en matière d'orientation notamment dans les Classes Préparatoires aux Grandes Ecoles (CPGE) demeurent liées à l'origine sociale. Les enfants d'origine sociale aisée ont très largement plus de chances de satisfaire au concours d'entrée dans les dispositifs qui préparent à l'entrée dans les Grandes Ecoles.

D'autres caractéristiques sociales sont aussi passées en revue. Il s'agit notamment du sexe, l'âge, le lieu de résidence.

Non seulement les filles ont massivement bénéficié de l'effet de la démocratisation universitaire comme l'affirment les statistiques du ministère français de l'enseignement supérieur et de la recherche -MESR (2012), devenant largement majoritaire, l'effet du genre leur est aussi favorable (Michaut, 2012).

Les filles seraient en effet plus prédisposées à réussir dans leurs études, du fait de leur socialisation différentielle (Duru-Bellat, 2004). Cette explication est partagée par d'autres chercheurs (Lahire, 1997 ; Frickey et Primon, 2002 ; Gruel et Thiphaine, 2004). Néanmoins d'autres travaux n'observent pas un impact réel de l'effet du genre. C'est le cas de l'étude qui a été réalisée sur les étudiants de la Faculté des Sciences de la Nature et de la Vie (SNV) de l'Université de Bourgogne (M'piayi, 1999) et qui conclut à une différence liée au genre non significative.

L'explication sociodémographique de la réussite universitaire devenue insuffisante, a dû être complétée par la prise en compte des acquisitions initiales comme en témoignent de nombreux travaux. Souvent évalué par qualité du baccalauréat (mention ou note obtenue) et la série, l'effet du passé scolaire produit des résultats très significatifs. C'est le cas des mentions « bien » et « très bien » qui sont plus efficaces que la mention « passable » (Michaut, 2000). Duru-Bellat (1995) et Michaut (2020) montrent le lien positif entre la note du baccalauréat et les chances de passage en deuxième année universitaire. Quant à la série, les titulaires d'un baccalauréat général seraient plus efficaces que les bacheliers technologiques (Michaut, 2000).

3- Situation de la réussite dans l'enseignement universitaire public du Congo

3.1- Indicateurs globaux

D'une façon générale, les taux de réussite sont assez faibles mais relativement comparables à ceux qui sont observés ailleurs notamment en France. En 2017, 43.7% d'étudiants ont pu valider leur année. A titre de comparaison, 79% l'ont fait dans le privé. Les données individuelles par établissement indiquent des taux très bas en droit, avec 19% de taux de succès. Le taux de réussite le plus élevé est détenu par l'Institut Supérieur de Gestion -ISG- (53%). La Faculté des Sciences Economiques et la Faculté des Lettres et des Sciences Humaines affichent respectivement 37 et 31%. Pour donner quelques exemples comparatifs avec le secteur privé, on relève un taux suffisamment faible à l'Institut Universitaire du Congo (25%) mais une bien meilleure efficacité à l'Ecole Supérieure de Technologie. On constate ainsi des disparités similaires au sein du public et dans le privé, même si globalement la réussite est largement plus importante dans le privé pour des raisons assez évidentes : des taux d'encadrement plus faibles et une implication plus importante des acteurs.

Dans le privé, une assiduité et une motivation plus marquées tant chez les enseignants qu'au niveau des étudiants. Pour les premiers, à cause de la régularité des émoluments et du contrôle de gestion sociale auquel ils font face, tandis que pour les seconds, c'est la nécessité de rentabiliser l'investissement (le coût important des études) qui leur fait pression.

3.2- Description statistique de la population observée à la FSE

Cette présentation est faite de manière à distinguer la composition selon les caractéristiques individuelles et les taux de réussite dans un premier temps et la structure croisée dans un second temps.

3.2.1- Structure de la population en fonction des caractéristiques individuelles des étudiants et des taux de réussite

La population observée à partir d'une enquête menée directement auprès des étudiants est constituée de 205 sujets répartis entre la Licence 3 (37.7%), le Master 1 (53%) et le Master 2 (9.27%). Le tableau n°1 ci-dessous montre que sur le plan socio-démographique, une répartition égale de la population selon le genre (52.2% de garçons contre 47.8% de filles). 12.2% de l'ensemble de ces étudiants vivent en couple, 12% avec enfants à charge.

La répartition selon l'origine sociale est fortement à l'avantage des employés ou ouvriers (41.5%) suivis des artisans, commerçants et chef d'entreprise (31.4%), les cadres et professions libérales intellectuelles supérieures représentent à peine un cinquième (18%), les sans emploi et les professions intermédiaires occupent la part restante (6.3 et 2.9%). Comme on pouvait s'y attendre, la moitié des sondés habitent chez leurs parents. Seuls 7.5% sont logés en résidence universitaire alors que 35.6% sont locataires, les 2.4% restants ne bénéficient d'aucun de ces 3 modes d'hébergement. 12% de l'ensemble de l'échantillon vivent en couple et 12% ont un enfant à charge.

Ces données sont détaillées dans le tableau 1, présenté à la page suivante.

Tableau n°1 : Structure selon les caractéristiques socio démographiques

	N	%
Sexe		
Masculin	107	52,20
Féminin	98	47,80
CSP du père		
Cadre et profession libérale intellectuelle supérieure	37	18,0
Artisan, commerçant, chef d'entreprise	64	31,2
Profession intermédiaire	6	2,9
Employé ou ouvrier	85	41,5
Sans emploi	13	6,3
Domicile		
Campus	15	7,3
Chez les parents	112	54,6
Loue	73	35,6
Vit chez soi sans louer	5	2,4
Vie en couple		
Oui	25	12,20
Non	180	87,80
Obtention d'enfants		
Oui	25	12,20
Non	180	87,80
Total	205	100,00

Source : Auteur à partir des données de l'enquête

La répartition en fonction des acquisitions initiales (tableau n°2 ci-dessous) montre une prédominance des étudiants titulaires du baccalauréat série D (65.85%), suivis de très loin par les bacheliers des séries C (16.10%) et technique (18.05%). Parmi eux, seuls 12.7% ont obtenu une mention « bien » ou « assez bien » au baccalauréat. La grande majorité est dépourvue de mention, ce qui correspond en réalité à la mention minimum (passable). Si la mention avait un effet très significatif sur la réussite, on devrait alors s'inquiéter de l'efficacité de l'enseignement universitaire, étant donnée la proportion importante d'étudiants sans mention qui serait davantage plus élevée dans la population effective totale en Licence 2 et 3. Il faut insister sur le fait que l'observation portée sur les 3 dernières années des cycles LM (Licence Master) exclut les étudiants de L1 et L2 qui n'ont pas accédé en L3 au moment de l'enquête et qui seraient davantage sans mention au baccalauréat et nous aurions fort à parier que c'est dans cette population recalée que la part des sans mention est plus conséquente. Près de 45% de l'ensemble de la population observée ont passé le baccalauréat dans la capitale contre 16.59% à Pointe-Noire et 38.53% sur le reste du territoire. Cette répartition est conforme au poids démographique des deux villes capitales.

Tableau n°2 : Structure de la population en fonction des acquisitions initiales

	N	%
Série du BAC		
C	33	16,10
D	135	65,85
Technique	37	18,05
Mention du BAC		
Passable	179	87,32
Assez bien ou bien	26	12,70
Département du BAC		
Brazzaville	92	44,88
Pointe Noire	34	16,59
Autres	79	38,5
Total	205	100,00

Source : Auteur à partir des données de l'enquête

En ce qui concerne le degré de réussite, il s'agissait de mesurer la qualité de leur parcours antérieur depuis la première année à la Faculté de Sciences Economiques. L'absence de notes d'examen de l'année en cours, du fait de la COVID-19, nous a conduit à élaborer un autre indicateur qualitatif de mesure de la réussite : le nombre total de sessions passées pour accéder à chacun de niveaux retenus dans cette étude. Cette démarche est explicitée dans la section suivante qui expose la méthodologie de la recherche. Il s'agit donc d'un indice composite d'évaluation de la réussite défini comme suit :

On construit d'abord RL2 (réussite dans le cycle allant de la première à la deuxième année de licence) tel que :

$$RL_2 = NL_1S_1 + NL_1S_2 + NL_2S_1 + NL_2S_2$$

Avec $NL_nS_m = \{1, 2, 3\}$ est le nombre de sessions passées pour valider le semestre m (1 ou 2) dans niveau L_n (Licence 1 ou 2).

Par conséquent $RL_2 = \{4, 5, 6, 7, \dots\}$, puisqu'il faut au moins 4 semestres pour valider une L2 (2 semestres en L1 et 2 semestres en L2).

Ensuite on construit la variable de la réussite (REUS) telle que

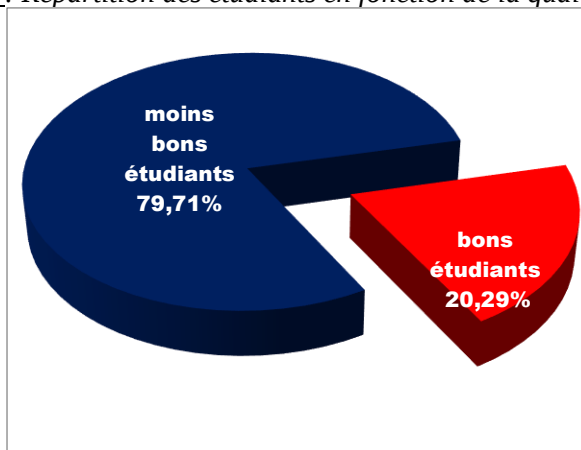
$$REUS = \begin{cases} 1 & \text{si } RL_2 = 4 \\ 0 & \text{si } RL_2 \geq 5 \end{cases}$$

Si $REUS_i = 0$ on considère que l'individu i est un « mauvais étudiant » ou « moins bon étudiant ». Et si, $REUS_i = 1$ on considère que l'individu i est un « bon étudiant ».

Comme on peut le lire sur le diagramme suivant, 79.71% d'étudiants sondés ont, à un moment donné de leurs cursus dans le premier cycle universitaire, éprouvé des difficultés. Seuls 20.29 sont considérés bons (score $RL_2=4$).

Les données statistiques de la réussite indiquent un taux de bons étudiants d'à peine 20%. Cela monte bien le caractère assez préoccupant de l'inefficacité de l'enseignement tout au moins dans l'établissement concerné par l'étude.

Diagramme n° 1 : Répartition des étudiants en fonction de la qualité de leur réussite



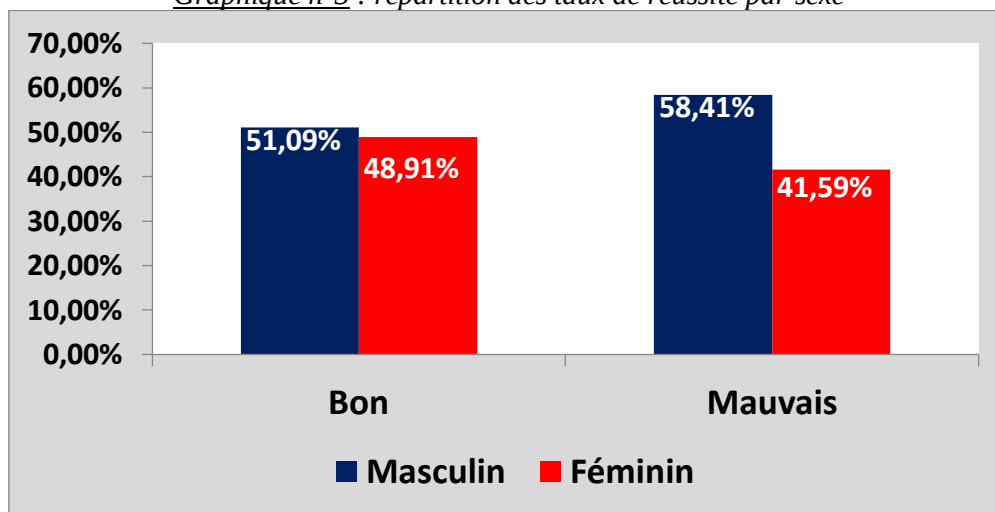
Source : Auteur à partir des résultats de l'enquête.

Nous pouvons dès à présent faire une première analyse de la réussite à partir des tris croisés ou approche statistique avant d'aborder l'approche économétrique.

3.2.1- Structure de la réussite en fonction des caractéristiques individuelles des étudiants

L'analyse du profil de la réussite à la Faculté des Sciences Economiques par sexe, comme le montre le diagramme ci-dessous, semble indiquer une différence assez négligeable entre les filles (48.9%) et les garçons (51.09%) en matière de réussite universitaire à la FSE de l'UMNG.

Graphique n°3 : répartition des taux de réussite par sexe



Source : auteur à partir des données de l'enquête

L'analyse de la réussite croisée avec les caractéristiques socio-démographiques des étudiants montre, d'après le tableau n°4 qui suit, que les taux de bons étudiants ont de meilleurs résultats chez les sujets de père cadre, profession libérale intellectuelle supérieure et chez ceux de père artisan commerçant et chef d'entreprise (24.91 et 23.65%). Les étudiants de père employé ou ouvrier et ceux de père sans emploi connaissent des taux nettement plus faibles, à peine un peu plus que 17% au mieux.

Au niveau familial, ce sont les étudiants vivant en couple qui paradoxalement présentent des meilleurs taux de réussite (31%) en comparaison avec leurs collègues hors couple (18.95%). De même, ce sont les étudiants ayant des enfants à charge qui paraissent plus efficaces, avec un taux de bonne réussite de 28.94% contre 18.04% pour ceux qui n'ont pas cette charge.

Il apparaît enfin que la vie en cité universitaire ne constitue pas un avantage en matière de réussite puisque ce sont les campusards qui enregistrent le taux le plus bas (7.8%). La résidence personnelle hors location serait la plus efficace (42.5%). Il n'y a pas de différence significative entre la résidence parentale et la location (21.26% et 20.30%).

Tableau n°4 : Réussite en fonction des caractéristiques socio-démographiques

	Réussite			P-valeur
	Mauvais	Bon	Total	
CSP du père				
Cadre et profession libérale intellectuelle supérieure	75,09%	24,91%	100,00%	0,006
Artisan, commerçant, chef d'entreprise	76,35%	23,65%	100,00%	
Profession intermédiaire	80,00%	20,00%	100,00%	
Employé ou ouvrier	83,54%	16,46%	100,00%	
Sans emploi	82,65%	17,35%	100,00%	
Domicile				
Campus	92,20%	7,80%	100,00%	0,000
Chez les parents	78,74%	21,26%	100,00%	
Loue	79,70%	20,30%	100,00%	
(autre) Vit chez soi	57,50%	42,50%	100,00%	
Vivre en couple				
Oui	68,42%	31,58%	100,00%	0,000
Non	81,05%	18,95%	100,00%	
Enfants à charge				
Oui	71,09%	28,91%	100,00%	0,001
Non	80,96%	19,04%	100,00%	
Total	79,71%	20,29%	100,00%	

Source : Auteur à partir des données de l'enquête

Cette décomposition socio-démographique laisse apparaître un phénomène constant, celui de l'échec universitaire au niveau de toutes les catégories, qui varie entre environ 60 et plus de 90%.

On fait le même constat lorsque la structure de la population croise cette fois-ci réussite et acquisitions initiales (tableau n°5 ci-dessous). Aucune catégorie ne se disculpe de la massification de l'échec dont le taux évolue entre presque 70% (bacheliers série D) et 90% (bacheliers du technique). La série C est celle qui protégerait le mieux contre l'échec. La mention obtenue au baccalauréat ne semble pas discriminante. 80% des étudiants, avec ou sans mention, auraient éprouvé des difficultés dans le parcours universitaire L1 à L2. Ce taux correspond d'ailleurs au taux global d'étudiants moins bons, présenté plus haut dans le diagramme. La répartition territoriale du passage du baccalauréat est, elle aussi, sans choc sur la réussite. On ne peut donc pas dire que les grandes villes préparent mieux les élèves que le reste du pays, à la poursuite des études à la FSE de l'UNMNG.

Tableau 5 : Réussite en fonction des acquisitions initiales

	Réussite			P-valeur
	Mauvais	Bon	Total	
Série du BAC				
C	68,28%	31,72%	100,00%	0,000
D	79,78%	20,22%	100,00%	
Technique	90,52%	9,48%	100,00%	
Mention du BAC				
Passable	79,64%	20,36%	100,00%	0,877
Assez bien ou bien	80,09%	19,91%	100,00%	
Département du BAC				
Brazzaville	80,39%	19,61%	100,00%	0,797
Pointe Noire	79,18%	20,82%	100,00%	
Autres	79,02%	20,98%	100,00%	
Total	79,71%	20,29%	100,00%	

Source : Auteur à partir des données de l'enquête

Les résultats selon le département du bac montrent qu'il n'y a pas de différence significative entre ceux qui ont obtenu le baccalauréat dans l'une ou l'autre des deux plus grandes villes du Congo et ceux qui l'ont obtenu dans le reste du pays.

4- Méthodologie

4.1- Les Hypothèses

Les très faibles taux de réussite signalent l'ampleur de l'échec universitaire. La littérature indique les pistes d'investigation explorées dans d'autres pays.

Nous pouvons donc considérer que les caractéristiques individuelles des étudiants (côté demande) et des établissements (coté offre) sont explicatives des différences de réussite entre les étudiants. La spécificité de la population observée (un seul établissement) fait que cet article analyse exclusivement l'impact des caractéristiques de la demande sur la réussite. Associée à la problématique posée en introduction, la revue de la littérature conduit aux hypothèses suivantes :

- (1) A propos des caractéristiques socio-démographiques à l'instar de l'origine sociale, on suppose par exemple que les étudiants de père ouvrier ou employé ont de moins bons résultats (plus faibles chances de valider les semestres que leurs homologues plus aisés), et que les garçons réussissent mieux que les filles ...
- (2) Pour ce qui est des acquisitions initiales et du parcours antérieur notamment scolaire, en se basant sur la série et la réussite au baccalauréat, on suppose que les étudiants qui ont obtenu les meilleures mentions réaliseraient de meilleurs résultats, et que le baccalauréat scientifique aurait un effet plus bénéfique que les autres séries (baccalauréats C et technologique).

(3)

4.2- Spécification des variables et du modèle

4.2.1- Modèle de base

Il convient d'élaborer une fonction de production qui mettrait en relation la réussite des étudiants et les variables caractéristiques des inputs utilisées dans le processus de production du bien éducatif et signalées dans les hypothèses de travail, en se servant du modèle logistique suivant développé par Guérieroux (1989) sur la modélisation économétrique des variables qualitatives et repris dans de nombreux travaux qui analysent les chances de survenance d'événements (cas présent des chances de réussite universitaire). En effet, le modèle de régression linéaire n'est plus approprié dans la mesure où, sa formulation conduirait à une équation avec deux membres de nature différente. L'un évalué sous forme de codes représentant les modalités de la variable qualitative, donc avec un ensemble de définition dénombrable. Le second membre, une combinaison linéaire de variables quantitatives et/ou qualitatives, pourrait prendre n'importe quelle valeur. Le principe dans ce cas consiste à modéliser la probabilité de survenance des différentes modalités et cela se fait généralement en utilisant une fonction de répartition.

Supposons une variable Y dépendante (réussite dans le cas de notre analyse) qui ne prend que les valeurs 0 (moins bons ou mauvais étudiant) ou 1 (bon étudiant). Pour un individu i de l'échantillon de taille n, Y prend la valeur Y(i). La base de données comporte j variables explicatives $X_1, X_2, \dots, X_j$ et pour un individu i, X(i) prend les valeurs $X_1(i), X_2(i), \dots, X_j(i)$.

Supposons que le fait d'être un bon étudiant est guidé par une variable Z non observée. Cette variable latente, qui par hypothèse s'adapte à une mesure quantitative décrit alors le niveau de la réussite. Ainsi, l'individu i est bon dès lors que Z(i) est supérieure à un certain seuil y_0 .

L'hypothèse émise sur la variable latente nous permet d'écrire d'une part

$$Z(i) = \alpha_0 + \left(\sum_{k=1}^j \alpha_k X_k(i) \right) + \varepsilon_i$$

Et d'autre part

$$Y = \begin{cases} 0 & \text{si } Z \leq y_0 \\ 1 & \text{si } Z > y_0 \end{cases}$$

α_0 : une constante

α_k : paramètres du modèle qui déterminent l'ampleur de l'effet de la variable X_k sur la réussite Y chez l'individu i.

De ce fait, la probabilité P_i que l'individu soit bon ($Y=1$) sera :

$$\begin{aligned} P_i &= P(Y = 1) = P(Z > y_0) = P\left(\alpha_0 + \sum_{k=1}^j \alpha_k X_k(i) + \varepsilon_i > y_0\right) \\ &= P\left[\varepsilon_i > y_0 - \alpha_0 - \left(\sum_{k=1}^j \alpha_k X_k(i)\right)\right] = P\left[\varepsilon_i \leq \alpha_0 + \left(\sum_{k=1}^j \alpha_k X_k(i)\right) - y_0\right] \\ A &= \Phi\left[\alpha_0 + \left(\sum_{k=1}^j \alpha_k X_k(i)\right) - y_0\right] \end{aligned}$$

Avec $\Phi(\cdot)$, la fonction de répartition de la loi de ε_i .

A défaut de disposer d'information sur la distribution de erreurs ε_i , on est amené à faire des hypothèses sur la fonction de répartition $\Phi(\cdot)$. On ferait ainsi appel au modèle logit, probit ou gombit selon que la fonction de répartition utilisée soit respectivement celle de la loi logistique, de la loi normale ou de la loi de Gumbel. Le modèle logit est le plus utilisé dans le domaine de l'éducation car il fait intervenir des Odds Ratio¹ (rapports des rapports des chances). Et lorsque la variable dépendante ne contient que deux modalités, on parle du modèle logistique binaire.

L'objectif du modèle est de construire une fonction qui permettra de prédire et expliquer les valeurs de la variable Y à partir de l'ensemble de descripteurs. Pour ce faire, la régression logistique binaire autorise l'hypothèse suivante :

Si $\Phi(\cdot)$ est la fonction de répartition de la loi logistique, alors en posant :

$$\begin{aligned} A &= \alpha_0 + \left(\sum_{k=1}^j \alpha_k X_k(i)\right) - y_0, \\ P_i &= P(\varepsilon_i \leq A) = \Phi(A) = \frac{e^A}{1 + e^A} = \frac{1}{1 + e^{-A}} \end{aligned}$$

Nous pouvons observer que si ε_i suit une loi logistique, alors sa fonction Logit s'écrit :

$$\ln\left(\frac{P_i}{1 - P_i}\right) = A = \alpha_0 + \left(\sum_{k=1}^j \alpha_k X_k(i)\right) - y_0$$

Le rapport $\frac{P_i}{1 - P_i}$ est appelé rapport de chances c'est-à-dire les chances de la survenance de l'évènement analysé au lieu de sa non survenance.

La méthode utilisée pour estimer les paramètres du modèle est celui du maximum de vraisemblance et la probabilité d'un individu est modélisée à l'aide de la loi binomiale. Ainsi la vraisemblance du modèle s'écrit :

$$L(p, X) = \prod_{i=1}^n \left[P_i^{y(i)} * (1 - P_i)^{1-y(i)} \right]$$

¹ Les odds ratio ou rapports des rapports de chances utilisés en sciences sociales mesurent la fluidité sociale c'est-à-dire la capacité de la société à ne pas faire dépendre la mobilité entre deux catégories, des forces particulières qu'exercerait chacune d'elle. C'est un indicateur de mesure d'inégalité. Un odds ratio évalué à 1 traduit une mobilité parfaite qui n'accuse aucune discrimination. Plus il s'éloigne de 1, plus la société structurée en ces deux catégories est inégalitaire.

La statistique de Wald permet de tester la significativité individuelle des variables, c'est-à-dire voir si chacune des variables influence significativement la variable dépendante. Les hypothèses dans ce cas sont les suivantes :

$$\begin{cases} H_0 : \alpha_k = 0 \\ H_1 : \alpha_k \neq 0 \end{cases}$$

Où α_k représente le coefficient associé à la variable explicative X_k .

Le calcul du R^2 de McFaden permet de mesurer la qualité d'ajustement du modèle.

Une fois que le modèle est estimé, il faut déterminer la qualité de l'ajustement du modèle aux données (le « Goodness of fit ») ou la capacité de prédiction.

Supposons par exemple les valeurs observées de la variable dépendante suivantes

$$y' = (y_1, y_2, \dots, y_n)$$

et les valeurs prédites par le modèle par

$$\hat{y}' = (\hat{y}_1, \hat{y}_2, \dots, \hat{y}_n)$$

où n est la taille de l'échantillon.

On considérera que le modèle est bon si :

- la distance entre la variable dépendante observée y' et la valeur prédite par le modèle $\hat{y}'$ est faible. Cette condition sera vérifiée à partir du test de Hosmer et Lemeshow ;
- le modèle prédit bien les valeurs $Y=0$ et les valeurs $Y=1$. C'est le tableau de classification qui instruit cette condition ;
- le modèle permet de bien discriminer entre les valeurs de $Y = 0$ et $Y = 1$ en fonction des variables explicatives $X_1, X_2, \dots, X_n$; autrement dit, on obtient de bonnes sensibilités, de bonnes spécificités et une bonne courbe ROC (Receiver Operating Characteristic ou caractéristiques opératoires de réception). Les démarches de ces différents tests seront exposées dans la section suivante.

Démarche dans l'interprétation des résultats

Les paramètres estimés à partir du modèle logistique permettent de calculer les Odds Ratio (OR), les écarts absolus des chances de survenance des événements, les chances de survenance des événements simple ou complexe (plusieurs critères).

Un OR est un rapport des cotes d'exposition chez les « individus positifs » (ou bons étudiants dans le cas de notre étude) et les « individus négatifs » (ou moins bons étudiants). C'est le rapport entre la probabilité de survenance d'un événement et de la probabilité de survenance de l'événement opposé, dans le groupe concerné. Par exemple supposons une seule variable explicative dichotomique comme le sexe, codée comme suit :

« 0 » pour féminin et « 1 » pour masculin, avec pour probabilités de survenance respectives de $P(Y = 0) = P_0$ et $P(Y = 1) = P_1$. Dans le cadre de cet exemple où l'on considère que seule la variable sexe est significativement explicative de la variable dépendante, la fonction logistique prendra la forme suivante :

$$\text{Logit}[P(Y = 1/\text{sexe})] = \alpha_0 + \alpha_1 * \text{sexe}$$

Ainsi, pour les hommes on aura :

$$\text{Logit}[P(Y = 1/\text{sexe} = 1)] = \alpha_0 + \alpha_1$$

Et pour les femmes, on aura :

$$\text{Logit}[P(Y = 1/\text{sexe} = 0)] = \alpha_0$$

Rappelons que α_0 et α_1 sont les coefficients du modèle logistique présenté plus haut, estimés en maximisant la fonction de vraisemblance associée.

Le Odds Ratio correspondant à la variable sexe sera :

$$OR = \frac{\frac{P_1}{1 - P_1}}{\frac{P_0}{1 - P_0}} = \frac{e^{(\alpha_0 + \alpha_1)}}{e^{\alpha_0}} = e^{\alpha_1}$$

La formule ci-dessus démontre que dans un modèle logistique, l'exponentielle du coefficient de la variable explicative s'interprète comme son Odds Ratio. Dans le cas où la variable explicative a plus de deux modalités et/ou est non ordonnée, on dichotomise les co-variables (toutes les modalités) sauf une qui sera considérée comme témoin ou variable de référence.

Lorsque le rapport est inférieur à 1 ($OR < 1$), c'est-à-dire :

$$\frac{P_1}{1 - P_1} < \frac{P_0}{1 - P_0}$$

On dira que la probabilité qu'un garçon soit bon étudiant plutôt que moins bon (mauvais) est plus élevée que celle observée chez une fille. L'inversion de l'inégalité inverse de facto l'interprétation. C'est la fille qui deviendrait plus efficace.

Si le $OR = 1$, il n'y aurait pas de discrimination fille/garçon car les chances pour un garçon bon étudiant de devenir moins bon sont exactement les mêmes que celles observées chez les filles. Dès lors, le fait d'être une fille ou un garçon n'affecte pas les chances d'être bon ou moins bon, donc sur la réussite. Dans ce cas, si différence de chances de réussite il y avait entre une fille et un garçon, elle ne serait pas liée au genre.

4.2.2 - Spécification des variables du modèle

Il convient de présenter avant tout la définition des variables, en distinguant la variable expliquée et les variables explicatives.

Rappelons qu'il s'agit dans ce travail d'identifier les facteurs explicatifs de la réussite universitaire à la Faculté des Sciences Economiques de l'Université Marien NGOUABI. C'est donc la réussite (REUS2) qui correspond à la variable endogène ou expliquée. Elle prend la forme d'un indicateur composite construit à cette fin de la manière qui suit, déjà exposée dans la section présentant les caractéristiques statistiques de la population. A titre de rappel, elle élaborée à partir de quatre variables intermédiaires (NL_1S_1 , NL_1S_2 , NL_2S_1 et NL_2S_2). Chacune d'elle a 3 modalités ; 1 (si l'étudiant a passé une seule session pour valider un semestre), 2 (si l'étudiant a passé deux sessions pour valider un semestre) et 3 (si l'étudiant a passé 3 sessions ou plus). De ces modalités on a construit un indicateur synthétique RL2 mesurant la qualité de la réussite et qui n'est autre que la somme des valeurs des quatre variables intermédiaires, avant de définir la variable REUS qui permet de classer les étudiants entre les « bons » et « les moins bons ou mauvais ».

$$REUS2 = \begin{cases} 1 & \text{si } RL2 = 4 \\ 0 & \text{si } RL2 \geq 5 \end{cases}$$

La valeur $REUS_i = 0$ signifie que l'individu i est un moins bon ou mauvais étudiant ;
 $REUS_i = 1$ signale un bon étudiant i .

En ce qui concerne les variables explicatives, elles sont structurées de sorte à distinguer les caractéristiques sociodémographiques et les variables mesurant le niveau initial. Pour ce qui est des variables sociodémographiques, nous en avons considéré cinq : le sexe, la CSP du père, le domicile, la vie en couple et les enfants à charge.

Le sexe permet d'évaluer l'inégalité fille/garçon, donc deux modalités : fille et garçon. Quant à la CSP du père qui définit l'origine sociale de l'étudiant, 4 modalités ont été retenues à savoir : cadre et profession libérale intellectuelle supérieure, artisan commerçant et chef d'entreprise, profession intermédiaire, l'ensemble ouvrier/employé et les sans emploi. La variable domicile a permis de distinguer les étudiants vivant en cité universitaire, ceux qui vivent chez leurs parents, ceux qui louent et enfin les autres. Il a aussi semblé judicieux de distinguer ceux qui vivent en couple et les étudiants parents, ayant à charge un ou plusieurs enfants.

Les variables mesurant les acquisitions initiales concernent essentiellement le baccalauréat, à travers sa série, la mention obtenue et le département de son passage. Nous avons distingué les séries C, D et les autres (technologique et professionnelle). La mention a permis de dissocier les lauréats passables (ou sans mention) de leurs collègues qui ont eu une distinction (bien, très bien voire plus). Trois modalités ont été considérées pour répartir les étudiants selon le lieu de passage du baccalauréat : Brazzaville, Pointe-Noire et le reste du Pays.

5 – Résultats économétriques des tests de robustesse

Il convient dans un premier temps d'identifier, donc de sélectionner les variables qui impactent la réussite universitaire, puis de déterminer le degré de robustesse du modèle à partir des tests exposés dans la section précédente avant de construire le modèle en dichotomisant lesdites variables pertinentes identifiées. Ce qui permettra par la suite de déterminer les écarts de chances de réussite entre les catégories d'étudiants, c'est-à-dire l'ampleur de l'impact de chaque modalité des variables sur la probabilité de réussite (« être bon ») des étudiants.

5-1 Identification des variables influençant la réussite

Les résultats du tableau ci-dessous révèlent qu'au seuil même de 10%, les variables sexe, enfants à charge, mention et département du baccalauréat, n'ont pas d'impact significatif sur la réussite universitaire. Toutes les autres variables (série du bac, la catégorie socioprofessionnelle du père, le domicile, la vie en couple) sont très influentes au seuil de 1%.

Tableau n°6 : résultats du modèle logistique de sélection des variables influentes

Iteration 0: log likelihood = -844.22399						
Iteration 1: log likelihood = -798.59894						
Iteration 2: log likelihood = -796.31461						
Iteration 3: log likelihood = -796.31017						
Iteration 4: log likelihood = -796.31017						
Logistic regression			Number of obs = 205			
			LR chi2(8) = 95.83			
			Prob > chi2 = 0.0000			
Log likelihood = -796.31017			Pseudo R2 = 0.0568			
REUSSITE1	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
SEXE	-.1845614	.1293624	-1.43	0.154	-.4381072	.0689843
CSP	-.1337011	.0275327	-4.86	0.000	-.1876641	-.0797381
DOMICILE	.3068104	.1005014	3.05	0.002	.1098312	.5037896
VIE_COUPLE	-.7617256	.2280641	-3.34	0.001	-1.208723	-.3147282
OBTENTION_ENFANT	-.0867473	.2196634	-0.39	0.693	-.5172798	.3437851
SERIE	-.5273501	.0842704	-6.26	0.000	-.6925169	-.3621832
MENTION	-.0558337	.1874854	-0.30	0.766	-.4232983	.311631
DEPARTEMENT	-.0481221	.0713613	-0.67	0.500	-.1879877	.0917436
_cons	2.043213	.6281332	3.25	0.001	.8120941	3.274331

Source : Auteur à partir des résultats du modèle

5.2- Les tests de robustesse

Comme nous l'avons annoncé, il convient de réaliser successivement les tests liés au classement et celui qui instruit le pouvoir discriminant du modèle.

5.2.1- Tableaux de classement du modèle

Commençons d'abord par présenter la démarche du test avant d'interpréter les résultats.

5.2.1.1- Présentation théorique

Rappelons d'abord que le modèle logistique permet de modéliser la probabilité des attributs 0/1 de la variable dépendante, notée Y, en fonction des co-variables (différentes modalités des variables) retenues. A partir des probabilités estimées, nous pouvons, en fixant un seuil de significativité, classer les individus dans la catégorie Y = 1 si sa probabilité est supérieure au seuil et dans la catégorie, Y = 0 dans le cas contraire. Il s'agit là d'une règle de classement.

Il est intéressant de déterminer la performance de classement et la façon dont celui-ci dépend de la règle choisie afin de voir s'il n'y a pas des observations très mal ajustées, c'est-à-dire dont le classement prédit par le modèle diffère de celui observé réellement, donc qui, de fait, perturberaient l'estimation des paramètres. Ce classement permet de déterminer le pouvoir

prédictif du modèle. Il doit être opéré préalablement à l'échelle globale de l'échantillon (**recherche du degré d'ajustement global**) puis, par la suite, séquentiellement à l'intérieur des différents sous-groupes préalablement définis à partir d'un critère adéquat comme les quantiles, de manière à opérer des regroupements des probabilité prédites par le modèle afin de voir s'il n'y a pas des sous-groupes mal ajustés qui auraient en conséquence un effet biaisant sur les estimateurs des paramètres dudit modèle. Dans ce dernier cas, il s'agit de l'évaluation de l'ajustement peaufiné signalant le degré d'adéquation au sein de chaque sous-groupe, donc la régularité d'ajustement dans l'ensemble de l'échantillon, appelée aussi **niveau de calibrage** du modèle. Un bon modèle doit déjà être globalement suffisamment ajusté et bien calibré.

Pour ce faire, la recherche du degré d'ajustement global soumet chaque individu « bon » ou « moins bon » ou « mauvais », au test de concordance entre sa probabilité d'être « bon » ou « moins bon » (donnée prédictive ou simulation) et son état de réussite effective (bon ou moins bon). Il convient dans ce cas de calculer les proportions d'étudiants déclarés « bons » (ou « moins bons ») par les simulations du modèle, dans le total d'étudiants effectivement « bons » (ou « moins bons »).

La recherche du degré de calibrage quant à lui consiste à calculer pour chacun des groupes, le nombre observé de réponses positives $Y = 1$ (étudiants bons) et négatives $Y = 0$ (étudiants moins bons) que l'on compare au nombre espéré (prédit ou simulé par le modèle), de manière à calculer par la suite la distance entre les fréquences observées et les fréquences simulées au moyen de la statistique khi-deux. Lorsque cette distance est faible (p-valeur supérieure au seuil de significativité) on considère que le modèle est bien calibré ou régulier dans son pouvoir prédictif. Pour cela, on fait recours au Test de Hosmer et Lemeshow.

5.2.1.2- Résultats des tests de classement

Les résultats du Test d'adéquation globale sont présentés dans le tableau n°7 ci-dessous. Ils indiquent que pour les bons étudiants (41 individus), la probabilité d'un individu de cette catégorie d'être effectivement bon est inférieure à 50%. En effet la concordance entre les chances d'être bon et l'effectivité de ces bons étudiants n'est vérifiée que pour 2.4% des cas au seuil de 5%. Ce qui est faible, puisque 97.6% d'étudiants bons sont mal ajustés par le modèle.

A l'inverse, pour les moins bons (164), la probabilité de concordance est supérieure à 50%. Elle est vérifiée pour 100% des cas au seuil de 5%. Aucun étudiant moins bon n'est déclaré bon par le modèle.

Même si, la capacité de prédiction des bons étudiants est très faible, on doit noter la parfaite prédiction des étudiants faibles. Ainsi, globalement, on peut constater, d'après l'ensemble de ces résultats, qu'au seuil de 5%, le modèle semble bien classer 165 étudiants (100% des 164 moins bons étudiants et 2.4% des 41 bons étudiants) sur les 205 observés. Ce qui fait un bon taux global de classement (80.5%). Le taux d'erreur est donc faible (19.5%).

Tableau n°7 : Qualité globale de prédiction du modèle

Logistic model for REUS			
Classified	True		Total
	D	~D	
+	1	0	1
-	40	164	204
Total	41	164	205
Classified + if predicted $\Pr(D) \geq .5$			
True D defined as REUSSITE1 != 0			
Sensitivity	Pr(+ D)	2.44%	
Specificity	Pr(- ~D)	100.00%	
Positive predictive value	Pr(D +)	100.00%	
Negative predictive value	Pr(~D -)	80.39%	
False + rate for true ~D	Pr(+ ~D)	0.00%	
False - rate for true D	Pr(- D)	97.56%	
False + rate for classified +	Pr(~D +)	0.00%	
False - rate for classified -	Pr(D -)	19.61%	
Correctly classified		80.49%	

Source : Auteur calcul à partir des données de l'enquête

En ce qui concerne la recherche du degré de calibrage du modèle via le test de Hosmer et Lemeshow, les résultats présentés dans le tableau n°8 qui suit montrent que le test est concluant, autrement dit la régularité de l'ajustement du modèle aux données, au sein de chacun des 10 groupes élaborés, est vérifiée. En effet, la valeur d'Hosmer-Lemeshow (5.63) est réduite et la p-valeur (58.39%) supérieure au seuil de significativité de 5%, règle par ailleurs observée parallèlement au sein de chaque groupe (p-valeur de chaque groupe supérieur à 5%).

Figure n°8 : Test de Hosmer-Lemeshow

lfit, group(10) table						
logistic model for REUS, goodness-of-fit test						
(Table collapsed on quantiles of estimated probabilities)						
Group	Prob	Obs_1	Exp_1	Obs_0	Exp_0	Total
1	0.5457	16.66268	12.9	10.52632	14.2	27.189
2	0.5785	6.136364	13.0	16.66268	9.8	22.79904
3	0.6321	6.136364	7.7	6.136364	4.5	12.27273
4	0.6620	12.27273	14.9	10.52632	7.9	22.79904
5	0.6742	37.71531	25.4	0	12.4	37.71531
6	0.6829	33.32536	37.1	21.05263	17.2	54.37799
7	0.6988	21.05263	14.7	1.42e-14	6.3	21.05263
8	0.7027	74.53349	64.1	16.66268	27.1	91.19617
9	0.7140	10.52632	15.0	10.52632	6.0	21.05263
10	0.7257	16.66268	16.5	6.136364	6.3	22.79904
number of observations =			205			
number of groups =			10			
Hosmer-Lemeshow chi2(7) =			5.63			
Prob > chi2 =			0.5839			

Source : Calculs de l'auteur sur à partir données de l'enquête

5.2.1- Détermination du pouvoir discriminant du modèle à partir de degré de sensibilité, de spécificité et de la courbe ROC

Le modèle gagne davantage en qualité lorsqu'en plus d'être globalement bien ajusté et suffisamment calibré, il est doté d'un pouvoir discriminant au sein des sujets sur la thématique analysée. On peut en effet avoir des cas de modèles effectivement bien ajustés, offrant un bon taux de classement, mais qui fournissent une mauvaise discrimination. Par conséquent, nonobstant le fait que notre modèle assure déjà un bon classement des étudiants bons et moins bons, il importe de vérifier sa capacité discriminatoire sur les deux catégories d'étudiants.

Pour le vérifier, nous allons mettre en application les procédures de sensibilité et spécificité qui débouchent sur la courbe ROC dont le principe est exposé comme suit.

Rappelons que le modèle logistique formalise la probabilité des attributs 0/1 de la variable dépendante Y en fonction des co-variables $X_1, X_2, \dots, X_n$. A partir des probabilités estimées, on décide en fixant un seuil, par exemple à 0.5, de classer l'individu dans la catégorie. $Y = 1$ si sa probabilité est supérieure au seuil et dans la catégorie $Y = 0$ sinon. Il s'agit d'une règle de classement.

Il est intéressant de déterminer la performance du classement et savoir comment celui-ci dépend du seuil (ou de la règle) choisi. Pour cela, nous allons considérer les notions de sensibilité et spécificité. La sensibilité est définie comme la probabilité pour le modèle de détecter un étudiant bon ($Y=1$), à un seuil donné. On peut ainsi faire varier cette probabilité en modifiant le risque d'erreur accepté (seuil). L'ensemble de points obtenus définit la sensibilité du modèle. Appliqué au cas des étudiants moins bons ou mauvais, ce raisonnement détermine la spécificité, c'est-à-dire la capacité du modèle à déterminer, à chaque seuil, les moins bons étudiants. La représentation graphique des points de sensibilité et de spécificité donne la courbe ROC (Receiver operating characteristic) dont l'éloignement de la droite d'égalité (droite de 45°), autrement dit l'aire mesurant cet espace reflète la capacité du modèle à discriminer les bons étudiants des moins bons ou mauvais...

Néanmoins, un modèle parfait aura une courbe ROC telle que l'aire qu'elle forme avec la droite de 45° mesure 1 unité. La surface de la courbe nous permet d'évaluer la précision du modèle dans sa capacité à discriminer les valeurs positives ($Y = 1$) des valeurs négatives ($Y=0$).

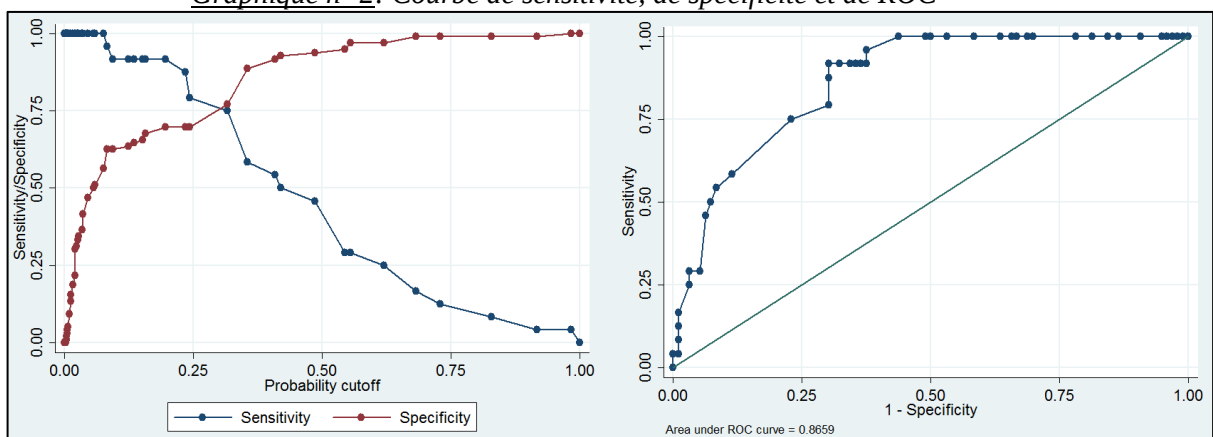
On retiendra la règle de décision suivante :

$\text{aire ROC} < 0,5$	alors, il n'y a pas de discrimination.
$0,5 \leq \text{aire ROC} < 0,7$	alors la discrimination est acceptable.
$0,7 \leq \text{aire ROC} < 0,9$	alors la discrimination est excellente.
$\text{aire ROC} \geq 0,9$	alors la discrimination est exceptionnelle.

La validation de toutes ces étapes garantit la bonne qualité du modèle élaboré.

Dans le cas notre étude, comme le montre le graphique n°1, en fixant le seuil à 0.30 on obtient un classement avec une sensibilité et une spécificité d'environ 75%. Ainsi, il ressort de du graphique que la discrimination est excellente car nous obtenons l'aire de ROC (0,8659) comprise entre 0.7 et 0.9.

Graphique n° 2: Courbe de sensibilité, de spécificité et de ROC



Source : Calculs de l'auteur à partir données de l'enquête

En guise de conclusion, nous pouvons finalement constater qu’au regard de la qualité des résultats des tests réalisés, le modèle initialement formalisé est robuste. Nous pouvons donc nous autoriser à analyser et interpréter les paramètres estimés.

5.3- Résultats du modèle dichotomisé, interprétation et implication en matière de politique éducative.

5.3.1- Résultats et interprétation

Le tableau n°9 présente les principaux résultats du modèle de régression élaboré. Ils signalent une résonance socio-démographique très classique qui veut que la dimension sociale favorise les enfants d’origine aisée. Les enfants dont le père est socialement mieux situé semblent effectivement plus performants que leurs collègues de père profession intermédiaire, employé ou ouvrier, et sans emploi. Ces derniers auraient respectivement 0.16 ; 0.24 ; et 0.22 unité de chances de réussite de moins que leurs collègues aisés, au seuil de 1%. Les Odds Ratio indiquent que les chances d’être un bon étudiant plutôt qu’un moins bon sont 3.23 fois moins élevées chez les étudiants de père profession intermédiaire, 2 fois moins chez ceux de père ouvrier ou employé et 2.2 fois moins chez leurs collègues de père sans emploi, par comparaison avec les étudiants d’origine plus ou moins aisés. Ce résultat témoigne du poids très pesant de l’origine sociale dans la réussite à l’UMNG et qui conduirait à la reproduction sociale.

Le domicile est aussi un facteur discriminant indiscutable au vu des estimateurs du modèle, avec une ampleur plus importante que celle de la tonalité sociale. Résider en cité universitaire induit un net désavantage par rapport aux autres modes d’hébergement suivants observés : résidence parentale, location et résidence personnelle hors location. On note un rapport de chances d’être un bon étudiant plutôt qu’un moins bon plus, défavorable aux campusards par rapport à chacune de ces trois autres catégories résidentielles de l’ordre de 3.23; 3.16 et 12 fois respectivement et qui, de fait bénéficient d’un écart de chances individuel moyen de 0.68¹; 0.66 et 0.7. On constate que la résidence hors cité universitaire, non parentale et non locative (autre résidence) est la modalité la plus efficace. On peut voir dans ces résultats, l’importance du cadre et du bon contexte des études qui ne semblent pas garantis par la résidence universitaire et /ou de la particularité de la catégorie d’étudiants concernés. Ce résultat, à première vue surprenant, incite à vouloir en savoir plus sur cette population et sur leurs conditions de vie.

Pour des raisons certainement similaires, la vie en couple semble plus bénéfique en termes de réussite universitaire. Un étudiant ne vivant pas en couple a 0.19 unité de chances de réussite de moins que son collègue qui vit en couple, au seuil de 1%. Ses chances d’être bon étudiant au lieu de moins bon ou mauvais sont 2.7 fois moins élevées que celles de l’étudiant vivant en couple.

1 L’écart de chances d’être bon étudiant, entre les résidents en cité universitaire et ceux qui vivent chez leurs parents est calculé comme suit :

$$\frac{1}{1 + e^{-(\alpha_0 + \alpha_1)}}$$

Avec α_0 la constante du modèle (-0.496) et α_1 le coefficient de la co-variable « résidence parentale » ce qui correspond à 0.68. Cela veut dire que sur 1 chance de d’être bon, un étudiant résidant chez ses parents a de ce fait 0.68 chance de plus qu’un campusard, autrement dit, ce dernier a 0.68 point de moins que son collègue. Il n’a donc que 0.32 chance.

Tableau 9 : Résultats du modèle dichotomisé

	Coef.	P-valeur	Odds ratio			Probabilité
			Valeur	Borne inf	Borne sup	
CSP du père : Cadre et profession libérale intellectuelle supérieure (Référence)						
Artisan, commerçant, chef d'entreprise et indépendant	-0,032	0.856	1,033	0,728	1,466	50,81%
Profession intermédiaire	-1,169	0.002	3,220	1,517	6,836	76,30%
Employé ou ouvrier	-0,661	0.000	1,936	1,368	1,106	65,94%
Sans emploi	-0,799	0.010	2,223	1,210	4,085	68,97%
Domicile : Campus (Référence)						
Chez les parents (résidence parentale)	1,171	0.001	0,3100	0,1581	0,6079	23,67%
Location hors campus	1,149	0.001	0,3171	0,1604	0,6266	24,07%
Autre résidence	2,485	0.000	0,0833	0,0330	0,2106	7,69%
Vie en couple ; Oui (Référence)						
Non	-1,004	0.000	2,729	1,863	3,996	73,18%
Série du bac ; C (Référence)						
D	-0,844	0.000	2,325	1,710	3,162	69,93%
Technique	-1,605	0.000	4,977	3,092	8,012	83,27%
Constante	- 0,496	0.313	1,642	0,224	1,614	62,15%

Source : Auteur à partir des résultats du modèle

Les acquisitions initiales confirment leur implication dans la réussite universitaire, comme on l'observe dans la plupart des études conduites dans le domaine. Cette analyse montre que ce sont les bacheliers de la formation technique qui ont les moins bonnes prédispositions acquisitionnelles conformes aux attentes universitaires en Sciences économiques, suivis des bacheliers série D. Les détenteurs du baccalauréat série C seraient donc les meilleurs, au seuil de 1%. Les écarts absolus des chances de réussite entre cette dernière catégorie et chacune des deux autres sont respectivement de 0.89 et 0.207 unité respectivement. Les chances d'être bon étudiant au lieu de moins bon, sont respectivement 5 et 2.4 fois moins importantes par rapport aux bacheliers de la série C. Ce résultat confirme l'idée largement répandue à propos du prestige du baccalauréat série C qui serait la crème de la formation pré-baccalauréat et de son impact souvent observé dans les analyses (Michaut, 2000).

5.2.1- Traduction des résultats en matière de gestion et de politique éducative

Il est toujours important et socialement utile de traduire les résultats des études en recommandations en matière décisionnelle tant pour les pouvoirs publics que pour les acteurs de la formation à leurs différents rangs. Le premier résultat qui interpelle est celui de la résidence estudiantine qui devrait logiquement être un facteur significatif de la réussite universitaire. Il apparaît donc indispensable de conduire une grande enquête sur les conditions de vie et d'études des étudiants concernés car ce résultat laisse soupçonner la dégradation de cet aspect. Si tel était le cas, on pourrait recommander l'inscription de la restauration des cités universitaires et l'élargissement du parc de logements au plus près des établissements universitaires dans les

priorités futures des chantiers de l'université, et plus globalement, la réhabilitation du service social qui aurait vocation à apporter l'aide nécessaire à l'étudiant(e) concerné(e). Une Maison de l'Etudiant devrait être installée sur le site universitaire à l'instar de ce qui se passe ailleurs afin que la prise en charge universitaire des étudiants ne se limite pas exclusivement à l'intégration abrupte des connaissances (dimension cognitive) mais prenne en compte l'étudiant dans toutes ses dimensions. Ce dernier n'est pas qu'un consommateur de connaissances mais un Être pluriel qui évolue dans un univers complexe. L'impact sociologique défavorable aux étudiants de classes populaires plaide pour l'instauration d'une vraie politique sociale protectrice des étudiants les plus vulnérables.

La deuxième recommandation est associée au poids des acquisitions initiales qui, fort heureusement, ne semble pas être porté par la mention obtenue au baccalauréat, même s'il s'agit d'un résultat assez controversé qui mérite un réexamen. Une étude de grande ampleur développant des outils d'évaluation des acquisitions initiales (recueil des scores obtenus au baccalauréat, organisation d'une session spéciale en début d'année afin de voir l'impact de ce niveau initial sur la réussite en fin d'année).

La correction du déficit cognitif qui affecte la réussite universitaire peut être réalisée par anticipation en amont, en offrant aux futurs étudiants, dès la classe de terminale, des outils d'une intégration facile et rapide dans l'enseignement supérieur (préparation à la pédagogie universitaire, révision des programmes scolaire, initiation à la recherche, échange avec les étudiants, et enseignants du supérieur, portes ouvertes avec possibilité d'assister aux cours, mise en place des mécanismes d'orientation...) ou en aval (principe du tutorat d'accompagnement, mise à niveau ou une refonte complète de la pédagogie universitaire de manière à l'adapter aux capacités des étudiants).

Conclusion

Le constat de la massification de l'échec universitaire a attiré notre attention de sorte à initier cette recherche qui s'est interrogée sur l'influence des facteurs socio-démographiques et des acquisitions initiales des étudiants sur leur parcours durant les deux premières années universitaires à la FSE de l'UMNG. A partir d'une enquête de terrain, une population d'étudiants composée des sujets ayant traversé les deux premières années de licence, donc inscrits en Licence 3 (L3) et en Master (M1 et M2), a été interrogée de manière à relever les informations ces différents aspects. La réussite définie par la qualité du parcours (4 sessions-semestres strictes passées ou au contraire plus de 4) distingue les bons étudiants des moins bons (ou mauvais) et a été reliée aux différentes caractéristiques des étudiants à travers une fonction de production du bien éducatif). Le traitement des données a nécessité une batterie des tests sophistiqués autorisant ainsi la spécification d'un modèle robuste d'estimation des paramètres. La nature qualitative de la variable dépendante « réussite universitaire » a imposé l'emploi de modèle logistique. Les estimateurs ont permis de calculer les écarts de réussite entre les catégories d'étudiants distinguées par les modalités des variables testées significatives et les rapports de chances (Odds Ratio) d'être bon étudiant au lieu de moins bon, entre les catégories.

Les résultats montrent le rôle très significatif de la série du baccalauréat général série C scientifique, à un moindre niveau, de celui du baccalauréat série D en référence au baccalauréat technique. De même, les étudiants qui vivent en résidence universitaires ont moins de chances

de réussite que leurs collègues qui sont domiciliés chez leurs parents ou ailleurs, en tant que locataires ou non. L'analyse de la vie en couple montre que celle-ci est un facteur de réussite. Quant à l'origine sociale, son effet classique plus favorable aux catégories aisées y est bien présent

Ces résultats signalent soit l'effet des prédispositions différenciées véhiculées par les contenus pédagogiques des programmes lycéens selon la série du baccalauréat, soit des apports sociaux et économiques qui déterminent les conditions de vie en fonction de la résidence, ou la particularité de la population concernée. Ces interprétations suggèrent quelques recommandations allant dans le sens de l'amélioration du cadre et des conditions de vie de l'étudiant en matière de résidence, mais aussi par une prise en charge plus globale en mettant en place un service social efficace. Une politique favorisant la bonne transition lycée-université est indispensable. Enfin, des études complémentaires plus poussées devraient être envisagées afin de creuser davantage certains aspects soulevés par cette recherche (conditions de vie des étudiants) ou non explorés à l'instar de la motivation ou le degré d'implication des étudiants dans leur fonction, leur jugement sur l'état de l'Université ou encore le rôle des facteurs institutionnels autrement dit l'effet de l'offre.

Bibliographie

- Bertrand C. (2014) « Soutenir la transformation pédagogique dans l'enseignement supérieur ». Paris: Ministère de l'Enseignement Supérieur et de la Recherche.
- Bocquillon M., Derobertmasure & A. Demeuse M. (2017) « Les recherches sur l'enseignement efficace en bref (2e édition) » in Working papers de l'INAS.)
- Boujut E. & Bruchon-Schweitzer M. (2007) « Rôle de certains facteurs psychosociaux dans la réussite universitaire d'étudiants de première année » – L'Orientation Scolaire et Professionnelle 36, 2 (157-177).
- Bressoux P. (1993). « Les effets des écoles et des classes sur l'apprentissage de la lecture ». Thèse de doctorat non publiée, Université de Bourgogne, Dijon.
- Braxton J.-M., Bray N. J. & Berger J.-B. (2000) « Faculty Teaching Skills and their Influence on the College Student Departure Process » Journal of College Student Development 41, 2 (215-227).
- Bressoux, P. (1993) « Les effets des écoles et des classes sur l'apprentissage de la lecture ». Thèse de doctorat non publiée, Université de Bourgogne, Institut de recherche en économie de l'éducation, Dijon
- Danner M. (2000) « A qui profite le tutorat mis en place dans le premier cycle universitaire ? » – Les Sciences de l'Éducation - Pour l'Ère Nouvelle 33, 1 (25-41).
- Duguet A. (2014) « Les pratiques pédagogiques en première année universitaire : description et analyse de leurs implications sur la scolarité des étudiants ». Dijon : thèse de doctorat, Université de Bourgogne.
- Duguet A., Lambert-Le Mener M. & Morlaix S (2016). Les déterminants de la réussite à l'université. Quels apports de la recherche en Éducation ? Quelles perspectives de recherche ?. Spiral-E. Revue de recherches en éducation, supplément électronique au n°57, :
- Duru-Bellat M. (1995) « Des tentatives de prédiction aux écueils de la prévention en matière d'échec en première année d'université » – Savoir 3 (399-416).
- Felouzis G (2000) & Felouzis G. (2000) « Repenser les inégalités à l'université. Des inégalités sociales aux inégalités locales dans trois disciplines universitaires » – Sociétés Contemporaines 38 (67-97).
- Felouzis G. (2003) « Les mutations actuelles de l'Université ». Paris : PUF.

- Felouzis G. & Le Guyader M. (2007) « Qu'enseigne-t-on en première année de licence ? La place de la recherche dans l'enseignement universitaire » – Revue de l'Inspection Générale 4 (71-77).
- Fortier M. S., Vallerand R. J. & Guay F. (1995) « Academic motivation and school performance : toward a structural model » – Contemporary Educational Psychology 20 (257-274).
- Fouquet S. (2013) « Parcours et réussite en licence et en master à l'université » – Note d'information de la DEPP 02/13.
- Frickey A. & Primon J.-L. (2002) « Les manières sexuées d'étudier en première année d'université » Sociétés Contemporaines 48 (63-85).
- Goux D. & Maurin, E. (1995). « Origine sociale et destinée scolaire. L'inégalité des chances à travers les enquêtes FQP 1970, 1977, 1985, 1993 » Revue française de sociologie, 36 (1), 81-121
- Gruel L. & Thiphaine B. (2004) « Des meilleures scolarités féminines aux meilleures carrières masculines ». Paris : Observatoire national de la vie étudiante.
- Gouriéroux, C. (1989). Économétrie des variables qualitatives. 2^{ème} édition, Paris : Economica.
- Hanushek, E.-A. (1979) «Conceptual and empirical issues in the estimation of educational production function» In J.-P. Jarrousse (1991) L'économie de l'éducation : du capital humain à l'évaluation des processus et des systèmes éducatifs. Perspectives documentaires en éducation, 23, 79-105.
- Lahire B. (1997) « Les Manières d'étudier ». Paris : La Documentation Française.
- Lemaire (2000) & Lemaire S. (2000) « Les facteurs de réussite dans les deux premières années d'enseignement supérieur (Deug, DUT, BTS) » Note d'information 0025.
- Menga-Mokembi R. I. (2019) « Efficacité de l'enseignement supérieur en République du Congo ». Thèse de doctorat non publiée. Université Marien NGOUABI
- Michaut C. (2000) « L'influence du contexte universitaire sur la réussite des étudiants. Dijon : Thèse de doctorat, Université de Bourgogne.
- M'piayi A. (2001) « Contribution à l'évaluation de l'efficacité des dispositifs de remédiation » Mesure et Evaluation en Education, volume n°22 pp. 45-67
- M'piayi A. (1999) « Les déterminants de la réussite universitaire : le cas du DEUG B à l'université de Bourgogne ». Thèse de doctorat non publiée, Université de bourgogne.
- MESR (2013) « L'état de l'enseignement supérieur et de la recherche en France ». Rapport.
- Nightingale P. & O'Neil M. (1994) « Achieving quality in learning in higher education ». Londres : Kogan.
- Pau J. J. (2007) « L'Economie de l'Education » Armand Colin.
- Psacharopoulos, G., & Woodhall, M. (1988). « L'éducation pour le développement. Une analyse des choix d'investissement ». Paris : Economica.