



HAL
open science

Identifier et comprendre les difficultés d'adaptation des primo entrantes à l'université: utilisation d'une méthode mixte quantitative-qualitative avec des méthodes statistiques d'apprentissage automatique

Patrick Pamphile, Isabelle Bournaud, Céline Clavel

► To cite this version:

Patrick Pamphile, Isabelle Bournaud, Céline Clavel. Identifier et comprendre les difficultés d'adaptation des primo entrantes à l'université: utilisation d'une méthode mixte quantitative-qualitative avec des méthodes statistiques d'apprentissage automatique. DIRES 2024 - Diversité, Réussites[s] dans l'Enseignement Supérieur, Apr 2024, Nantes (France), France. hal-04489836v1

HAL Id: hal-04489836

<https://hal.science/hal-04489836v1>

Submitted on 5 Mar 2024 (v1), last revised 9 Apr 2024 (v2)

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.

Proposition de communication individuelle

« Identifier et comprendre les difficultés d'adaptation des primo entrantes à l'université : utilisation d'une méthode mixte quantitative-qualitative avec des méthodes statistiques d'apprentissage automatique »

Auteurs :

Pamphile Patrick, Laboratoire de Mathématiques d'Orsay-CNRS, Inria CELESTE

patrick.pamphile@universite-paris-saclay.fr

Université Paris-Saclay, 91405 Orsay France

Bournaud Isabelle, UR Études sur les Sciences et Techniques ,

isabelle.bournaud@universite-paris-saclay.fr

Université Paris-Saclay, 91405 Orsay France

Clavel Céline, Laboratoire Interdisciplinaire des Sciences du Numérique- CNRS,

celine.clavel@universite-paris-saclay.fr

Université Paris-Saclay, 91405 Orsay France

Mots-clés : *Transition Lycée-Université ; Primo-entrantes ; Adaptation académique ; Méthodes mixtes ; Apprentissages statistiques ; Apprentissages automatiques*

Type de texte soumis (au choix)

- Texte court (500 mots, hors références)
- Texte long (2000 mots, hors références)

Type de communication

- Présentation d'une recherche
- Présentation d'un dispositif
- Retour d'expérience sur une initiative

Dans cette communication, il a été choisi de ne pas utiliser l'écriture inclusive pour des questions d'accessibilité du texte, et de faire primer le féminin sur le masculin car les personnes concernées par les propos sont en majorité de genre féminin.

1. Introduction

La transition lycée-université est considérée comme un moment critique pour les primo-entrantes qui doivent s'adapter non seulement à un environnement académique différent, mais aussi à de nouvelles interactions sociales et à une plus grande responsabilité et autonomie (Paivandi, 2019). L'intégration des primo-entrantes est ainsi une préoccupation majeure pour les universités, en particulier pour les Instituts Universitaires de Technologie (IUT). En effet, en plus de la réforme du Baccalauréat général en 2021, les IUT ont été confrontés à l'augmentation de 30% à 50% des quotas d'admission de bachelières technologiques, ce qui a eu un impact sur la diversité des profils des primo-entrantes.

Pour accueillir cette diversité et favoriser la réussite et l'épanouissement de ces étudiantes, il est essentiel de comprendre leurs besoins et leurs attentes, et de recueillir leurs impressions concernant leur expérience. Pour ce faire, nous avons adopté une méthode mixte combinant une enquête quantitative et une enquête qualitative auprès de primo-entrantes en IUT (Creswell & Clark, 2017). Les enquêtes quantitatives sont efficaces pour détecter des tendances générales, mais peuvent manquer de profondeur en ne permettant pas une expression ouverte. D'un autre côté, les enquêtes qualitatives génèrent des données textuelles riches qui permettent de comprendre de manière approfondie et individualisée la situation étudiée. Cependant, le traitement de ces données est plus complexe, ce qui tend à limiter la quantité de données exploitées. Les avancées techniques en apprentissage automatique offrent désormais la possibilité de traiter rapidement d'énormes volumes de données, quantitatives ou textuelles, permettant ainsi de tirer parti de l'articulation de méthodes quantitatives et qualitatives.

2. Méthode

Nous avons recueilli des données auprès d'étudiantes de première année de deux formations de BUT¹ d'un IUT d'Ile de France :

- des données quantitatives via la version française du *Student Adaptation to College Questionnaire* (SACQ) (Carayon et Gilles, 2005 ; Larose et al., 1996). Le SACQ évalue divers aspects de l'expérience étudiante, notamment l'adaptation académique, sociale et personnelle. Cette approche multidimensionnelle permet de capturer différents aspects de l'adaptation des étudiantes à la vie universitaire et de questionner leur satisfaction à l'égard des services universitaires.
- des données qualitatives via des rapports d'étonnement rédigés en fin de 1^{er} semestre de formation. Le rapport d'étonnement vise à capturer les impressions, attentes et difficultés potentielles des étudiantes lors de la transition du lycée à l'université. Son objectif principal est de stimuler une réflexion critique chez l'étudiante, l'incitant à évaluer de manière approfondie sa première expérience à l'université.

¹ BUT : Bachelor Universitaire de Technologie

2.1. Analyse des données quantitatives

La version française courte du SACQ comprend 41 questions qui portent sur différents aspects de l'adaptation de l'étudiante : son adaptation académique, ses attentes par rapport à la formation, son intégration sociale et d'éventuelles difficultés dans la gestion de sa vie quotidienne.

Le questionnaire a été administré en ligne au printemps 2023 ; 343 étudiantes ont répondu. Pour chaque proposition, les étudiantes étaient invitées à se positionner sur une proposition visant à évaluer leur capacité d'adaptation sur les différents thèmes, en utilisant une échelle de Lickert à 5 valeurs. Nous avons traité les données du SACQ à l'aide d'analyses statistiques non supervisées (Murphy, 2022).

2.1.1. Analyse Factorielle Exploratoire

Une Analyse Factorielle Exploratoire a permis d'identifier les corrélations latentes entre les différentes variables du questionnaire. Nous avons retenu un modèle à cinq groupes de variables corrélées correspondant à différentes dimensions de l'adaptation des primo-entrantes (voir Tableau 1).

Tableau 1 : Dimensions de l'adaptation obtenues à partir de l'analyse factorielle des réponses au SACQ

Questions concernant	Exemples de question	Nom attribué à la dimension
La formation	« La formation que je suis correspond à mes attentes »	Adaptation à la Formation (AF)
L'ajustement académique	« Je rencontre des difficultés à suivre le rythme des enseignements »	Ajustement Académique (AA)
Les interactions avec les pairs	« Je me suis fait des amies dans ma formation »	Interactions avec les Pairs (IP)
La gestion du quotidien	« Je rencontre des difficultés à gérer mon quotidien (repas, RDV médicaux, démarches administratives...) »	Adaptation Personnelle (AP)
Les interactions avec les enseignantes	« À la rentrée, je me suis sentie accueillie par l'équipe pédagogique »	Interactions avec les Enseignantes (IE)

2.1.2. Clustering

Ensuite, nous avons utilisé un algorithme de clustering des individus en se basant sur leurs réponses au SACQ, ce qui a conduit à l'identification de trois profils d'adaptation (voir section Résultats).

2.2. Analyse des données qualitatives

En fin de premier semestre 2022, les étudiantes ont été invitées à rédiger un rapport d'étonnement de 600 à 700 mots, reflétant leurs étonnements lors de leur arrivée à l'IUT et sur leur formation (Thievenaz, 2017). Ces rapports constituent les verbatims de notre étude.

Nous avons réalisé une analyse de contenu de ces verbatims afin d'identifier les sujets abordés dans l'ensemble des verbatims, puis à caractériser chaque verbatim à partir des sujets identifiés. L'objectif de cette analyse de contenu est d'affiner la caractérisation des profils d'adaptation obtenus lors du clustering des répondants au SACQ.

Pour automatiser ce traitement des verbatims, nous utilisons des outils de traitement du langage naturel (Church, 2017 ; Hardeniya & al. 2016). Les algorithmes de clustering ou de classification se basant sur des données numériques, une première étape d'encodage est nécessaire. Elle consiste à convertir les données textuelles en vecteurs numériques. Pour cela, nous avons opté pour l'utilisation d'un réseau neuronal à deux couches car ils sont plus sensibles à la sémantique que l'encodage basé sur la fréquence des mots, tout en étant moins complexes et plus faciles à mettre en œuvre que les réseaux neuronaux profonds.

L'analyse du contenu des verbatims est alors faite en deux étapes.

2.2.1. Identification des sujets abordés dans l'ensemble des verbatims

Les verbatims ont été découpés en phrases, et chaque phrase a été encodée en un vecteur numérique. Ces phrases ont été regroupées en 57 clusters à l'aide d'un algorithme de *k-means*. Chaque cluster de phrases correspond à un sujet abordé par les étudiantes. Les sujets ont été nommés en se basant sur les mots les plus fréquents dans chaque cluster et en se référant aux dimensions identifiées lors de l'analyse factorielle (voir Tableau 1).

Par exemple, le Cluster1 regroupe des phrases dans lesquelles les étudiantes expriment leur déception concernant leurs résultats académiques : « J'ai eu une note de 11,5/20 alors que je m'attendais à avoir plus de 15/20. », Ce cluster s'inscrit dans le thème Ajustement Académique : nous l'avons nommé **AA_Insatisfaction_Resultats**.

Le Cluster 57 regroupe des phrases dans lesquelles les étudiantes expriment des difficultés à trouver un équilibre entre les études et la vie personnelle : « J'ai eu de moins en moins de temps pour moi, pour passer du temps avec ma famille et mes amis, et cela m'a beaucoup affecté psychologiquement. », Ce cluster s'inscrit dans le thème Adaptation Personnelle : nous l'avons nommé **AP_Impact_Vie_Personnelle**.

2.2.2. Classification des verbatims

Pour chaque verbatim, nous avons calculé la probabilité que le verbatim aborde chacun des 57 sujets identifiés (voir un extrait dans le Tableau 2).

Tableau 2 : Les dix premiers verbatims et probabilités de classification des six premiers sujets (par ordre alphabétique).

Num_verbatim	AA_Difficulte_Adaptation	AA_Difficulte_Comprehension	AA_Difficulte_Concentration	AA_Difficulte_Motivation	AA_Difficulte_Organisation	AA_Difficulte_Rythme
V1	0,0159	0,0061	0,0045	0,0172	0,0230	0,0121
V2	0,0169	0,0090	0,0034	0,0256	0,0223	0,0114
V3	0,0151	0,0079	0,0048	0,0187	0,0208	0,0127
V4	0,0189	0,0065	0,0051	0,0173	0,0198	0,0074
V5	0,0110	0,0074	0,0044	0,0136	0,0194	0,0176
V6	0,0156	0,0063	0,0100	0,0348	0,0255	0,0115
V7	0,0164	0,0093	0,0036	0,0177	0,0185	0,0097
V8	0,0109	0,0102	0,0056	0,0111	0,0202	0,0144
V9	0,0133	0,0076	0,0040	0,0176	0,0184	0,0078
V10	0,0123	0,0081	0,0055	0,0132	0,0211	0,0166

3. Résultats

L'analyse des données quantitatives obtenues par le SACQ nous a permis d'identifier cinq dimensions de l'adaptation des primo-entrantes, ainsi que trois profils distincts d'étudiantes en matière d'adaptation. Pour caractériser ces profils, nous avons comparé les moyennes des différents clusters pour chacune des cinq dimensions. Avec une p-value <0,01, on peut conclure que :

- **le Cluster1** regroupe des individus ayant, en moyenne, des valeurs inférieures à la moyenne de l'ensemble des étudiantes pour les cinq dimensions. Ce profil correspond donc à des étudiantes qui rencontrent potentiellement des difficultés d'adaptation dans les cinq dimensions.

L'analyse du contenu des verbatims des individus de ce cluster met en évidence que ces étudiantes mentionnent plus fréquemment que les autres :

- le fait d'avoir choisi la formation pour son aspect professionnalisant,
- leur insatisfaction de l'aspect professionnalisant de la formation,
- leurs difficultés à s'organiser efficacement pour réussir,
- leurs difficultés à maintenir une concentration efficace,
- leur interrogation sur les différences entre les départements GEA1 et GEA2 au sein de l'IUT de Sceaux, en particulier sur les modalités de contrôle des connaissances.

Par ailleurs, les étudiantes de ce cluster abordent moins fréquemment que les autres la nécessité d'avoir à faire face à une charge de travail intense.

- **le Cluster2** regroupe des individus ayant, en moyenne, des valeurs inférieures à la moyenne de l'ensemble des étudiantes pour les dimensions : Adaptation Personnelle, Interactions avec les Pairs et Interactions avec les Enseignantes. Les étudiantes de ce profil déclarent des difficultés pour gérer le quotidien de la vie étudiante en dehors du cadre universitaire et dans leur intégration sociale. En revanche, on constate qu'elles ne rencontrent potentiellement pas de difficultés dans les dimensions Ajustement Académique et Adaptation à la Formation.

L'analyse du contenu des verbatims des individus de ce cluster montre que ces étudiantes mentionnent plus fréquemment que les autres :

- leurs difficultés d'adaptation à leur nouvel environnement et en particulier des différences et ajustements nécessaires lorsqu'elles passent du lycée à l'IUT, que ce soit en matière de structures académiques (Cours Magistraux et Travaux Dirigés) et d'attentes en termes d'apprentissage (rétention à long terme à l'IUT par rapport à l'apprentissage axé sur les contrôles au lycée) et environnement d'étude (temps de transport plus long),
- la difficulté à concilier études, loisirs et vie personnelle,
- en revanche, une expérience positive sur l'accueil et l'intégration à l'IUT.

- **le Cluster3** regroupe des individus ayant, en moyenne, des valeurs supérieures à la moyenne de l'ensemble des étudiantes sur les cinq dimensions. Les étudiantes de ce profil déclarent ne pas rencontrer de difficultés d'adaptation à l'IUT.

L'analyse du contenu des verbatims des individus de ce cluster montre que ces étudiantes mentionnent plus fréquemment que les autres l'importance des projets et travaux en équipe, qui :

- renforcent les liens entre les membres de l'équipe et favorisent la cohésion ;
- favorisent l'apprentissage en groupe et l'acquisition de compétences sociales et professionnelles.

Par ailleurs, les étudiantes de ce cluster mentionnent moins fréquemment que les autres :

- leurs difficultés d'adaptation à leur nouvel environnement ;

- leurs difficultés à s'organiser efficacement pour réussir.

4. Discussion

Alors que l'enquête quantitative a permis de révéler des profils différents d'adaptation à l'IUT, l'analyse du contenu des rapports d'étonnement a permis de qualifier plus précisément les difficultés rencontrées par les individus de ces différents profils. Cette qualification permet d'élaborer des solutions spécifiques pour faciliter l'adaptation des primo-entrantes à l'IUT.

A partir de l'expérience des étudiantes du cluster1, l'équipe pédagogique de la formation peut veiller à :

- améliorer la communication auprès des étudiantes sur l'aspect professionnalisant de la formation ;
- aider les étudiantes à développer des compétences d'organisation et de gestion du temps ;
- aider les étudiantes à améliorer leur capacité à se concentrer et à maintenir leur attention sur leurs tâches académiques.

A partir de l'expérience des étudiantes du cluster2, l'équipe pédagogique peut veiller à :

- améliorer l'accueil à l'IUT ;
- mettre l'accent sur les attentes des enseignantes sur le travail à faire en CM et en TD, les attentes en matière d'apprentissage et les éventuelles difficultés liées à l'environnement d'étude (temps de transport plus long).

A partir de l'expérience des étudiantes du cluster3, l'équipe pédagogique peut veiller à :

- aider les étudiantes à développer des compétences d'organisation et de gestion du temps,
- promouvoir les projets et travaux en équipe.

L'adoption d'une approche mixte combinant des méthodes quantitatives et qualitatives, dont les résultats sont traités et analysés avec des méthodes statistiques d'apprentissage automatique, offre la perspective d'acquérir une compréhension holistique des obstacles et besoins des étudiantes primo-entrantes. Cette approche englobe tous les aspects de leur adaptation à l'université, qu'ils soient académiques ou personnels. Ainsi, les équipes pédagogiques peuvent non seulement répondre aux difficultés et besoins des étudiantes pour faciliter l'adaptation à l'université et la réussite de toutes, mais aussi tenir compte des besoins individuels spécifiques de chacune afin de leur offrir une expérience universitaire positive.

Références

Carayon, S. et Gilles, P. Y. (2005). Développement du questionnaire d'adaptation des étudiants à l'université (QAEU). *L'orientation scolaire et professionnelle*, (34/2), 165-189.

Church, K. W. (2017). Word2Vec. *Natural Language Engineering*, 23(1), 155-162.

Creswell, J. W., & Clark, V. L. P. (2017). *Designing and conducting mixed methods research*. Sage publications.

Hardeniya, N., Perkins, J., Chopra, D., Joshi, N., & Mathur, I. (2016). *Natural language processing: python and NLTK*. Packt Publishing Ltd.

Larose, S., Soucy, N., Bernier, A. et Roy, R. (1996). Exploration des qualités psychométriques de la version française du Student Adaptation to College Questionnaire. *Mesure et évaluation en éducation*, 19(1), 69-94.

Murphy, K. P. (2022). *Machine learning: a probabilistic perspective*. MIT press.

Paivandi, S. (2019). Le défi de la transition entre secondaire et supérieur. Construisons des ponts. *Rapports du CNESEO*. https://www.cnesco.fr/wp-content/uploads/2019/08/190801_Cnesco_Post-baccalaureat_Paivandi.pdf

Thievenaz, J. (2017). *De l'étonnement à l'apprentissage : enquêter pour mieux comprendre*. De Boeck Supérieur.