



ERUDITE

**ÉQUIPE DE RECHERCHE SUR L'UTILISATION
DES DONNÉES INDIVIDUELLES EN LIEN
AVEC LA THÉORIE ÉCONOMIQUE**

Sous la co-tutelle de :
UPEC • UNIVERSITÉ PARIS-EST CRÉTEIL
UPEM • UNIVERSITÉ PARIS-EST MARNE-LA-VALLÉE

Series of ERUDITE Working Papers

N° 16-2019

Title

Négocier les règles d'évaluation
pour développer l'accompagnement à la réussite des étudiants

Author

Rémi LE GALL

Négocier les règles d'évaluation pour développer l'accompagnement à la réussite des étudiants

Rémi LE GALL¹

ERUDITE (EA 437) Université Paris-Est Marne-la-Vallée ; Université Paris-Est Créteil
TEPP (FR 3435)
CEET

Mai 2019

¹ Contact : 5, Boulevard Descartes – Cité Descartes – Champs-sur-Marne 77454 Marne-la-Vallée
CEDEX 2. remi.legall@u-pem.fr.

Résumé

Nous avons impliqué des étudiants de licence dans la co-construction de la règle d'évaluation de leur participation pendant les travaux dirigés afin d'améliorer leur réussite à l'université. Nous avons ainsi testé l'effet de la négociation contractuelle du dispositif de vérification de la participation à partir de la mesure des notes aux examens et de leur probabilité de valider leur semestre.

Le choix de ce terrain de recherche permet d'étendre la portée de la relation d'agence à l'interaction d'instruction. L'enseignant utilise l'évaluation des connaissances pour certifier le niveau d'acquisition de l'étudiant et l'inciter à réaliser un effort d'apprentissage. Cependant, à l'université, les modalités de tests n'ont pas été adaptées à un contexte de démocratisation de l'accès à l'enseignement supérieur et d'accueil de bacheliers de plus en plus nombreux. De plus, l'université fait face à un problème régulier d'échec des étudiants, notamment en première année de licence. Dans ce cadre, nous envisageons l'innovation dans les règles d'évaluation des étudiants comme un élément d'amélioration de leur réussite dans leur parcours de formation.

Nous partons ainsi du postulat que dans le cadre de la relation enseignant-étudiant, le dispositif de vérification de l'effort d'acquisition de connaissances et de compétences peut être négocié afin de parvenir à une situation socialement optimale, en réallouant plus efficacement les ressources de l'éducation à des tâches plus efficaces. Nous montrons que ce dispositif expérimental produit un impact positif sur la réussite à l'université dans certaines conditions, notamment en fonction de la manière dont est exécuté le contrat d'évaluation négocié.

Mots clés : relation d'agence ; contrat ; dispositif de vérification de l'information ; évaluation formative ; réussite à l'université ; expérimentation contrôlée de terrain avec assignation aléatoire.

Abstract

We involved undergraduate students in the co-construction of the evaluation rule of their participation during the tutorials in order to improve their success at the university. We thus tested the effect of the contractual negotiation of the participation verification device from the measurement of exam scores and their probability of validating their semester.

The choice of this research field makes it possible to extend the scope of the agency relationship to the instructional interaction. The teacher uses knowledge assessment to certify the student's level of achievement and motivate them to make a learning effort. However, at university, the test procedures have not been adapted to a context of democratization of access to higher education and of receiving more and more graduates. In addition, the university faces a regular problem of student failure, especially in the first year of the bachelor's degree. In this context, we conceive innovation in student assessment rules as an element of improving their success in their training course.

We thus start from the postulate that in the context of the teacher-student relationship, the verification mechanism of the effort to acquire knowledge and skills can be negotiated in order to reach a socially optimal situation, by reallocating resources more efficiently from education to more effective tasks. We show that this experimental device has a positive impact on university success under certain conditions, including the way in which the negotiated evaluation contract is executed.

Keywords: agency relationship; contract; information verification device; formative evaluation; success at the university; controlled field experiment with random assignment.

JEL : D2 ; D86 ; C93 ; I21 ; I23

1. INTRODUCTION

En 2017, l'indicateur de réussite à l'université, calculé par le Ministère de l'Enseignement supérieur, mesurait que parmi les néo-bacheliers inscrits en 2013-2014, moins de trois étudiants sur dix décrochaient leur licence en trois ans (Razafindratsima, 2018). En tenant compte des trajectoires de réorientation ou de redoublement, seuls quatre étudiants sur dix de cette même promotion ont obtenu leur licence en quatre années. Depuis l'entrée en vigueur de l'arrêté du 23 avril 2002, relatif aux études universitaires conduisant au grade de licence, cet indicateur est stable et il ne semble pas connaître d'inflexion à la hausse (voir annexe a).

Alors que le taux d'échec était considéré aussi bien comme une sélection qui s'opérait naturellement que comme un gage d'excellence du diplôme délivré et un signe de l'échec personnel des étudiants à réussir à l'université (pour une revue de la littérature récente des caractéristiques sociodémographiques et les profils de l'échec, se reporter à Aina et al., 2018, Duguet et al., 2016), la massification de l'enseignement supérieur a redéfini son interprétation. Ce phénomène pose désormais la question de la qualité de l'organisation du cursus des étudiants des universités qui paraissent ainsi accuser un retard dans l'adaptation des parcours pour accompagner le flux de ces « nouveaux publics »².

Pour rendre compte des tensions qui existent dans les trajectoires de formation et dans les modalités d'enseignement spécifique à l'université, il nous semble opportun de mobiliser la relation enseignant-étudiant dans les établissements de l'enseignement supérieur (Correa et Gruver, 1987). Sous plusieurs hypothèses, cette relation peut se caractériser comme une relation d'agence entre un enseignant (le principal) et des étudiants (les agents) au sein de laquelle les pratiques d'évaluation des connaissances jouent le rôle de dispositifs de vérification de l'information, c'est-à-dire de contrôle de l'effort d'apprentissage fourni par les étudiants (Zubrickas, 2016).

² Nous pourrions néanmoins nuancer cet argumentaire en présentant la multitude de dispositifs de remédiation à l'œuvre à l'université, tant au niveau national qu'au niveau local. Le « Plan Réussite en Licence », doté de 730 millions d'euros sur la période 2007-2012 a reformulé la licence pour promouvoir une spécialisation progressive sur les trois années de la préparation du diplôme et un renforcement de l'encadrement pédagogique grâce à davantage d'heures hebdomadaires pour accompagner les étudiants. En 2013, le « Plan National Vie Étudiante » a mis l'accent sur les conditions de vie des étudiants en développant l'accès au logement, à la santé et en réduisant les démarches administratives. Enfin, pour la période 2017-2022, le « Plan Étudiants » prévoit une meilleure lisibilité des formations ainsi qu'un approfondissement de l'orientation et de l'accès des lycéens à l'enseignement supérieur.

Au niveau local, le développement du tutorat qui vise à apporter une forme d'aide individualisée en dehors du contexte de la classe, la création de composantes de formation à la méthodologie pour favoriser le développement de compétences complémentaires à celles acquises pendant la formation et l'investissement dans des modules de renforcement des acquis liés à la discipline étudiée sont autant d'innovations qui ont tenté de favoriser l'accueil des étudiants. Dans le contexte français, il existe peu d'évaluations d'impact de ces dispositifs sur l'échec à l'université. À notre connaissance, une seule étude a estimé l'incidence du « Plan Réussite en Licence » sur les étudiants de l'Université de Bourgogne (Morlaix et Perret, 2013). Tout en relevant les difficultés de mesures de l'effet causal des dispositifs d'accompagnement, à cause de leur nombre élevé sur le terrain et de la méthodologie employée, les auteurs ont mesuré un effet nul ou non significatif des dispositifs. Les chercheurs de l'Institut de Recherche sur l'Éducation (IREDU, EA7318) notent que c'est finalement la diversité des opérations qui conduisent à brouiller la mesure des effets sans pouvoir imputer de lien de causalité propre à un dispositif particulier (accueil et suivi pédagogique renforcé, méthodologie de travail universitaire, réorientation des étudiants en situation d'échec ...).

Les rares recherches sur les dispositifs d'accompagnement méthodologiques ou pédagogiques montrent que le plus souvent ils sont facultatifs et touchent rarement ceux qui en auraient le plus besoin (Morlaix et Lambert-Le Mener, 2015).

En effet, dans cette relation, l'évaluation est un dispositif que les enseignants utilisent pour collecter des preuves du niveau d'apprentissage atteint par les étudiants. Il mesure ce que les étudiants connaissent, ce qu'ils comprennent et ce qu'ils sont capables de réaliser afin de produire une inférence sur l'apprentissage. C'est à la fois un outil de jugement sur les performances des étudiants, sur leurs forces et sur leurs faiblesses et de contrôle de leurs progrès par rapport à un but fixé. C'est un instrument d'aide à la notation et à la certification du niveau atteint par les étudiants pour la délivrance du diplôme (Bloom et al. 1971 ; McMillan, 2015 ; OCDE, 2013).

L'évaluation est également un dispositif qui permet d'inciter à l'apprentissage des étudiants en donnant les informations sur les attentes de l'enseignant ou de l'institution, en ce qui concerne l'acquisition d'un niveau de connaissances, de compétences et de savoir-faire. C'est un outil qui conditionne l'apprentissage et la motivation à apprendre en développant les compétences de planification et d'autonomie (Gibbs et Simpson, 2002).

La création d'un examen est à la discrétion de l'enseignant, qui est le seul concepteur et utilisateur des critères, des règles et des pondérations de l'évaluation. Ce rôle primordial lui confère une autorité et un pouvoir de contrôle sur l'apprentissage des étudiants, à l'instar de ce qu'il se réalise au sein d'une relation d'agence générique (Hölmstrom, 1979).

Mais cette dimension de la profession entre en conflit avec d'autres fonctions qui lui incombent : la recherche, l'enseignement, la responsabilité de formations, les activités professionnelles en parallèle de la formation. Il existe donc un arbitrage dans le temps dévolu à l'évaluation. Cet arbitrage conduit à choisir des modalités d'évaluation les plus rapides possible, qui demandent un temps de jugement le plus court possible, afin d'attribuer du temps à des tâches davantage valorisées. En d'autres termes, le coût d'opportunité de l'évaluation est élevé. Paradoxalement, la valeur attribuée à cette tâche est négativement liée au temps nécessaire pour réaliser l'évaluation. En effet, à l'évaluation s'ajoute la période de correction et le nombre d'étudiants à évaluer augmente tendanciellement.

Par conséquent, les formes d'évaluation à l'œuvre à l'université sont davantage de nature à opérer un classement des étudiants afin de pouvoir sélectionner ceux qui seront les plus capables de continuer leur formation. Cette utilisation de l'évaluation a pour finalité de contrôler l'apprentissage, ce que les chercheurs spécialistes de l'évaluation scolaire ont dénommé par l'évaluation « *summative* ».

Cependant, cette manière d'évaluer produirait un effet « *backwash* », c'est-à-dire qu'elle créerait un phénomène contraire à ce qu'elle est censée produire (OCDE, 2013). L'évaluation « *summative* » provoquerait un processus qui s'autoalimente : la peur de se retrouver dans les statistiques de l'échec à l'université déclenche des stratégies qui entrent en conflit avec ce qui serait la meilleure stratégie pour contrer l'échec. Alors que l'évaluation traditionnelle de l'université devrait permettre de valider l'acquisition des compétences spécifiques à une économie de la connaissance à savoir le développement de l'apprentissage tout au long de la vie et la capacité à être autonome pour résoudre des problèmes inconnus (Birenbaum, 1996), elle renforce d'autres compétences qui produisent de l'échec : les étudiants ne développent pas les compétences attendues et échouent aux examens dédiés à valider l'acquisition de l'autonomie, de l'adaptation, de la prise d'initiative, etc. Elle provoque une distanciation des étudiants avec la construction d'un goût, d'une appétence ou d'une motivation intrinsèque pour l'apprentissage, qui renforcerait la curiosité et l'esprit critique sur l'environnement de l'étudiant. Par conséquent, elle provoque une

dépossession de l'apprentissage, c'est-à-dire que les étudiants n'apprennent pas pour l'acquisition et la conservation durable des connaissances et des compétences, mais uniquement pour réussir à un examen. Malgré ces conséquences négatives, l'évaluation resterait résistante au changement et ce serait le dispositif le plus conservateur de l'université (Bloxham, 2016).

Pourtant, une seconde forme d'évaluation a été conceptualisée pour améliorer les pratiques et dépasser ces effets négatifs. L'évaluation « *formative* » est dédiée à la promotion de l'apprentissage (Scriven, 1967 ; Brown et Knight 1994 ; Black et William, 1998). L'évaluation pour l'apprentissage intègre le retour sur le résultat, qui se définit comme la production d'un commentaire sur le résultat à un examen de connaissances. Elle montre où se situent les erreurs et donne les moyens de remédiation. Cette pratique se conclut généralement par un nouvel examen pour apprécier les pistes d'amélioration³.

L'évaluation « formative » favorise également la transparence de l'évaluation. Elle intègre l'explicitation précise de l'articulation du critère d'évaluation et les objectifs à atteindre. Un lien clair et explicite entre ce qui est enseigné et ce qui est évalué permet aux étudiants de se focaliser sur des points importants de l'enseignement. Cela permet de s'assurer que la charge de travail est réaliste et que l'évaluation n'est pas menaçante. Par conséquent, cela renforce l'autorégulation de son propre travail, c'est-à-dire qu'elle rend capable de se situer soi-même dans l'acquisition des connaissances. Le concept de littéracie de l'évaluation traduit cette idée (Rust et al., 2003). Cela permet d'augmenter leurs compétences d'évaluation des pairs et d'autoévaluation. In fine, l'évaluation pour l'apprentissage permet également de former les étudiants à maîtriser la capacité d'évaluer des preuves, de juger des situations et les circonstances, de tirer des conclusions et d'agir en conséquence.

Il apparaît nécessaire de trouver un équilibre dans le paramétrage de l'évaluation entre sa dimension de mesure de performance et celle de promotion de l'apprentissage. En termes organisationnels, il s'agit de développer un nouveau contrat optimal d'évaluation des connaissances dans la relation étudiant-enseignant de l'enseignement supérieur. Dans ce contexte, nous souhaitons déterminer dans quelle mesure une relation enseignant-étudiant dans laquelle réside davantage de réciprocité, où chacun a un rôle et une voix pour influencer la participation dans l'évaluation permettrait-elle de favoriser la réussite des étudiants ? Dans ce cadre, il s'agit de rompre la balance du pouvoir d'évaluation dans cette relation d'agence particulière, où l'agent, en l'occurrence l'étudiant, aurait plus de contrôle, de choix et de flexibilité quant aux modalités de ses examens d'acquisition de connaissances. La conception de l'évaluation qui peut être configurée afin d'impliquer les deux parties de la relation, d'ouvrir à la négociation et de mettre le curseur aussi bien sur le processus que sur le résultat permettrait de renforcer des compétences qui conduisent les étudiants à améliorer leurs résultats à l'université.

Cette question permet d'approfondir les connaissances sur les interactions individuelles et les réactions comportementales liées aux dispositifs de vérification de l'information au sein des organisations. En effet, bien qu'ils soient destinés à collecter, enregistrer et analyser de l'information pour réduire l'aléa moral lié à une incertitude sur le comportement des agents, en les disciplinant (Pratt et Zeckhauser, 1985), ils seraient également capables de renverser la

³ Le « *feedback* » est un des rares concepts généraux en sciences de l'éducation clairement validés par les chercheurs (McKeachie, 1962). Derrière cette idée, l'évaluation doit servir à informer de façon pertinente les apprenants aussi bien qu'elle permet d'informer les autres parties prenantes des résultats de l'évaluation.

motivation intrinsèque à réaliser les tâches confiées (Frey, 1993). Il est donc opportun d'ouvrir la boîte noire des systèmes de surveillance (Gillet et al., 2016).

Nous avons tenté de répondre à cette question en expérimentant une forme nouvelle d'évaluation qui passe par la négociation et la co-construction des critères de notation de la participation des étudiants au sein de deux modules de formation de la licence d'Économie-Gestion de l'Université Paris-Est Marne-la-Vallée. Cette expérimentation exploratoire a permis de faire dialoguer un chargé de travaux dirigés avec des groupes d'étudiants sur ce qu'il est pertinent de considérer quand il s'agit d'examiner et donner un score au niveau de la participation dans la formation. Nous mesurons l'impact de ce dispositif sur leurs résultats aux examens de fin de semestre ainsi que sur la probabilité de valider le semestre.

Dans un premier temps, nous réaliserons une revue de la littérature de l'économie de l'éducation pour montrer qu'il existe des outils conceptuels qui permettent de justifier de l'utilisation de la négociation de l'évaluation dans la relation enseignant-étudiant, ainsi que des études en sciences de l'éducation qui ont commencé à produire des résultats novateurs sur le partenariat dans l'évaluation. Dans un deuxième temps, nous présenterons le protocole expérimental et le recueil de données. Enfin, les résultats de cette étude montreront que la participation des étudiants à la construction des modalités d'évaluation joue un rôle substantiel sur la réussite des étudiants.

2. REVUE DE LA LITTÉRATURE

a. Théorique

En économie de l'éducation, la théorie du capital humain procure un cadre unifié des décisions des étudiants à investir dans l'éducation, la durée et le type de scolarité et de formation et indique comment ces investissements déterminent leur productivité et leur revenu. Ce cadre théorique est la base conceptuelle de nombreuses recherches sur les incitations à rester à l'école, sur le retour sur investissement dans des années de scolarité supplémentaires, sur la qualité des formations, sur le choix de la formation professionnelle, sur les variations du salaire, et sur les liens avec le marché du travail (Card, 1999).

Un cadre alternatif s'emploie pour l'analyse économique des décisions prises au sein des écoles et des classes, c'est-à-dire le processus d'enseignement et d'éducation, et sur les relations entre les individus qui agissent dans ces organisations. En effet, à l'instar de l'analyse des décisions individuelles prises par les ménages ou les firmes, la microéconomie de l'instruction examine l'organisation générale et l'allocation des ressources au sein des structures d'enseignement (Brown et Saks, 1981 ; Brown et Saks, 1987).

Dans ce cadre d'analyse, l'enseignement ou l'instruction (« *schooling* ») correspond à un processus dans lequel les préférences de l'étudiant et les préférences de l'enseignant sont associées avec d'autres ressources pour former l'apprentissage. C'est « la courbe de l'apprentissage »⁴. Dans la

⁴ Elle équivaut formellement à la fonction de production de la firme, parce qu'elle correspond à l'idée qu'il faut associer des intrants, c'est-à-dire des facteurs d'apprentissage, qui sont transformés d'une certaine manière pour produire un bien, en l'occurrence le résultat de l'apprentissage. Traditionnellement, la courbe d'apprentissage est considérée comme étant non linéaire et conforme à la « loi des rendements décroissants », c'est-à-dire que le temps passé pour étudier procure un gain marginal de plus en plus faible en termes de réussite.

relation étudiant-enseignant, chaque intervenant joue ainsi un rôle actif pour la production de l'apprentissage.

L'étudiant a un comportement dual (Bishop, 2006), c'est-à-dire qu'on le considère comme un consommateur / investisseur qui choisit de répartir un temps optimal pour l'apprentissage et comme un producteur de l'apprentissage, qui reçoit les instructions, les conseils, et les évaluations de ses enseignants. Il possède un éventail de choix entre fournir un effort maximal pour l'apprentissage et l'abandon complet de la formation.

L'enseignant joue également un rôle dual. C'est un consommateur / investisseur et sa décision d'enseignement dépend de ses options extérieures et un producteur à part entière de l'apprentissage, ce qui lui permet d'être récompensé par l'effort fourni en transmettant des connaissances. Dans ce modèle, l'instruction associe dans la même dimension le contrôle et la délivrance des connaissances.

Dans cette relation, chaque individu choisit le niveau d'effort optimal pour l'apprentissage qui maximise son utilité étant donné la contrainte d'effort pour l'apprentissage pour acquérir (délivrer) des connaissances pour l'étudiant (enseignant), le niveau de récompense perçu par l'apprentissage, et la courbe d'apprentissage.

Le résultat lié à l'apprentissage se définit comme un indicateur de la réussite des étudiants, ou a minima l'indicateur du niveau de l'étudiant après être passé dans l'instruction et il est assimilable à un bien public.

Brown et Saks (1981) ainsi que Correa et Gruver (1986) montrent au travers de ces modèles que les préférences individuelles semblent être le principal déterminant de l'effort et du niveau de l'apprentissage, ce que nous interprétons comme le goût pour l'apprentissage. L'appétence pour les études pour les étudiants ou le goût pour la transmission de connaissances pour l'enseignant seraient les uniques déterminants de la réussite. Or, il n'y a pas d'éléments supplémentaires qui permettent de penser que le niveau estimé d'effort de l'autre agent est celui qui est véritablement réalisé. Cette situation conduit donc à des réajustements dans les comportements des deux parties.

En tenant compte des interactions entre les individus qui se réalisent au sein de la relation, c'est-à-dire que les niveaux d'effort d'apprentissage et d'instruction de l'étudiant et de l'enseignant se modifient en réaction à ce que décide l'autre individu, Correa et Gruver (1986) ont démontré l'existence d'un équilibre stable, mais qui peut être Pareto dominé par une situation d'interaction avec coopération sur le niveau d'effort réalisé dans l'apprentissage par l'enseignant et l'étudiant⁵. C'est un cas spécifique obtenu dans la littérature de l'allocation des ressources pour la production d'un bien public. En effet, les théoriciens admettent généralement que des optimisations indépendantes conduisent à des résultats qui sont inférieurs à l'optimum de Pareto. La raison pour laquelle l'interaction sans coopération conduit à des résultats scolaires inférieurs est que chaque individu considère les bénéfices pour eux-mêmes et pas les externalités qui bénéficient aux autres.

Selon les auteurs, une stratégie de menace serait susceptible d'opérer un déplacement suffisamment incitatif vers une solution supérieure à une situation sans coopération. Dans ce cas, il est légitime de penser que des interactions entre l'étudiant et l'enseignant qui permettraient de dialoguer sur les attentes réciproques sur le niveau d'effort et sur les dispositifs de contrôle de

⁵ L'équilibre est stable si et seulement si la valeur accordée à l'effort d'apprentissage par l'étudiant, c.-à-d. sa préférence, est « proche » de la valeur accordée à la réussite scolaire, et que la valeur accordée à l'effort d'instruction par l'enseignant est « proche » de la valeur accordée à la réussite scolaire.

connaissances seraient favorables à l'atteinte de résultats scolaires plus importants qu'en l'absence de négociation. Sans davantage de précision sur les termes de la négociation, il apparaît donc possible d'améliorer la réussite des étudiants en faisant dialoguer les deux parties de la relation.

L'approfondissement de ces interactions qui permet de tenir compte des modalités d'évaluation se situe dans un cadre contractuel, qui précise davantage les termes de la relation étudiant-enseignant. Ces modèles partent de l'hypothèse que l'éducation n'est plus un bien public, mais qu'il procure des bénéfices individuels. Par conséquent, des divergences d'intérêt existent à propos ce que procure ce bien.

Pour aligner les intérêts des deux parties, il faut résoudre un problème d'incitation. Dans ce cadre, le dispositif d'évaluation des connaissances, qui tient un double rôle : mesurer le niveau de connaissance et inciter à l'effort d'apprentissage.

Partant de cette idée, Zubrickas (2015) énonce qu'il existe une vision dichotomique de l'éducation : d'une part, elle assure la formation du capital humain et d'autre part elle signale les compétences acquises sur le marché du travail. Les enseignants sont concernés par la première vision, tandis que les étudiants s'intéressent au signal. Dans ce contexte, les étudiants traitent la connaissance comme quelque chose d'abstrait, d'éphémère et de coûteux, comparativement aux opportunités extérieures. C'est ce type d'attitude qui crée un conflit d'intérêts entre l'enseignant et l'étudiant. Dans cette relation, l'enseignant fait donc face à un problème de sélection des étudiants qui auraient des intérêts alignés avec les siens.

L'enseignant possède alors d'un dispositif, l'examen, qui lui permet d'évaluer le niveau de connaissances atteint par les étudiants. Par extension, l'enseignant définit une règle de notation qui assigne des scores au résultat de l'examen, c'est-à-dire au niveau de connaissances. Par hypothèse, l'examen est « parfait », c'est-à-dire qu'il y a une correspondance parfaite sur le niveau de connaissance évalué et la note donnée. La conséquence de ce postulat est que les notes assurent uniquement l'évaluation de la performance. Cette hypothèse permet de simplifier la modélisation, cependant elle empêche la formulation d'une description plus précise du contrôle des connaissances.

Ce modèle permet de qualifier le risque d'aléa moral sur les attentes des étudiants sur l'enseignement, au même titre qu'un aléa moral se manifeste dans les attentes des enseignants sur l'effort des étudiants. Le principal apport de ce type d'analyse est que ces divergences d'intérêt conduisent à une méfiance sur les résultats de l'éducation. Considérant le faible taux de réussite à l'université, il est permis de penser que les attentes de l'enseignement supérieur ne coïncident pas avec celles des étudiants. De surcroît, le contrat conclu entre les deux parties amènerait à des résultats inefficaces. Nous postulons qu'un des paramètres de ce contrat, à savoir le contrôle des connaissances, n'est pas efficacement calibré pour permettre d'atteindre un équilibre socialement partagé.

Pour apprécier l'effet de l'évaluation sur les comportements de la relation, il faut se situer au niveau de la théorie des standards d'éducation (Becker et Rosen, 1992 ; Michaelis et Schwanebeck, 2016). Dans ces modèles, la relation d'agence n'est pas explicitement évoquée, mais elle est sous-jacente. Un évaluateur cherche un examen qui maximise une fonction de bien-être social, sous-entendu le niveau de connaissance (le capital humain), en tenant compte de l'aversion à l'effort d'apprentissage des étudiants. L'analyse se concentre sur l'arbitrage entre l'effet d'incitation et de classement qu'un changement dans les modalités d'examen peut produire sur le comportement des étudiants.

Hormis, Michaelis et Schwanebeck (2016) qui s'intéressent à la réaction comportementale des étudiants à différentes variantes d'un examen de rattrapage (avec droit d'inscription, niveau requis au rattrapage différent de la première session, temps passé entre chaque examen, point malus), il ne semble pas exister d'autres tentatives pour examiner les méthodes d'évaluation à l'intérieur des classes, qui sont des manières plus régulières d'examiner l'apprentissage des étudiants que les examens nationaux⁶. Dans leur modélisation, l'évaluation conditionne l'effort d'apprentissage des étudiants, mais elle peut également causer des effets négatifs sur la réussite. Il s'agit alors de trouver un critère qui permettrait de désigner ce qu'il est souhaitable d'atteindre socialement.

Ce critère social peut être trouvé parmi les apports de la négociation dans le cadre de la théorie des contrats. En effet, une hypothèse traditionnelle de la relation d'agence est que le principal offre un contrat « à prendre ou à laisser », c'est-à-dire que le principal possède tout le pouvoir de négociation. Pourtant, dans des situations concrètes, les deux parties détiennent des marges de manœuvre pour négocier les termes du contrat.

Pour Pitchford (1998), ces situations recouvrent les cas où la responsabilité de l'agent est limitée, par exemple pour le remboursement des emprunts quand la capacité de paiement est faible. Dans ce cas, le pouvoir de négociation est représenté par la variation de l'utilité de réserve de l'agent, c'est-à-dire en faisant varier ses opportunités extérieures. Balckenborg (2001) utilise une approche différente en recourant à la solution de négociation à la Nash (Nash, 1950). Demougin et Helm (2004) étudient un jeu d'offre alternative avec aléa moral à la Rubinstein (1982). Enfin, Dittrich et Städter (2015) emploient la solution de négociation à la Kalai-Smorodinsky (Kalai et Smorodinsky, 1975).

La littérature sur la négociation dans un cadre d'aléa moral suggère qu'une courbe des contrats serait modélisable comme une frontière Pareto optimale où se situent les utilités du principal et de l'agent, à chaque fois qu'intervient une distribution du pouvoir de négociation entre les deux parties.

En revanche, le choix d'une solution théorique de négociation ne doit pas être une décision arbitraire puisque la préférence pour l'une ou l'autre solution peut sensiblement modifier les résultats de la négociation (Demougin et Helm, 2004). Dans le cas du changement de l'utilité de réserve et dans le jeu d'offre alternative, il n'y a pas d'effet sur la variété de choix du contrat optimal, c.-à-d. la courbe des contrats Pareto optimale n'est pas modifiée. Dans le cas de l'utilisation des solutions de négociation à la Nash ou à la Kalai-Smorodinsky, la courbe des contrats peut être modifiée.

Ce que nous retirons comme principal apport de cette littérature est qu'en présence d'aléa moral l'agent peut détenir un pouvoir de négociation, qui permet d'atteindre un optimum social plus efficace qu'en situation sans négociation.

Ce résultat est donc similaire à ce qui a été déterminé dans le cadre canonique de l'économie de l'instruction. Il est donc raisonnable de penser que des marges de manœuvre existent pour configurer une relation d'apprentissage où l'étudiant entrerait en négociation avec l'enseignant. Les termes de la négociation peuvent porter sur la règle d'examen, sur les notes attribuées lorsque l'examen est validé ou pas, sur le niveau d'effort attendu de l'étudiant pour acquérir des connaissances.

⁶ Rosen et Becker (1992), les règles d'évaluations permettent de modifier les résultats obtenus, mais il n'y a pas de spécifications sur les réactions comportementales individuelles. De plus, leur analyse se positionne au niveau d'un examen national, comme le baccalauréat.

Il est possible d'énoncer que dans le cadre d'une relation étudiant-enseignant, la négociation de la règle d'examen pour quantifier le niveau de connaissance obtenu serait une façon d'améliorer les positions des deux individus, et de contribuer à l'élévation du niveau de réussite des étudiants, y compris dans le cas où les deux parties éprouvent des divergences d'intérêt sur l'éducation.

b. Appliqué

En microéconomie appliquée, à moins d'une erreur de diagnostic, d'une part il ne semble pas exister de recherche qui porte sur la négociation du dispositif de vérification de l'effort dans un cadre de relation d'agence, et d'autre part il n'existerait pas de travaux qui se focalisent explicitement sur les pratiques d'examen des connaissances et leurs effets sur l'effort d'apprentissage⁷.

En revanche, dans la littérature ce sont les chercheurs en sciences de l'éducation qui traitent de cette question en déterminant dans quelle mesure la négociation des modalités de l'évaluation des connaissances permettait de dépasser les faibles résultats des étudiants, notamment à l'université. Les méthodes d'investigation employées sont à mi-chemin entre l'étude de cas et la recherche-action en recourant à des dispositifs d'analyse qualitatifs.

Le concept clé de cette partie de la littérature est celui de partenariat, conçu comme « un processus collaboratif et de réciprocité dans lequel tous les participants ont l'opportunité de contribuer équitablement, mais dans des conditions différentes, au programme ou à la conception pédagogique, à la prise de décision, à l'investigation et à l'analyse » (Bovill, et al., 2011). Cette pratique du partenariat permet de changer les rôles : l'enseignant substitue à sa position supérieure de surveillant une position d'accompagnant dans l'apprentissage ; l'étudiant troque sa fonction de consommateur passif du savoir à celle de contributeur actif à l'apprentissage.

C'est une collaboration dans le sens où les deux parties travaillent pour un objectif partagé : développer l'apprentissage par la compréhension et la création des objectifs et des critères d'évaluation. L'idée est de renforcer des compétences de contrôle et de jugement tout en développant une compréhension et un intérêt pour le processus d'examen des connaissances. Par conséquent, cela augmente la littéracie de l'évaluation et les étudiants sont capables de s'approprier l'évaluation (Rust et al., 2003). In fine, l'objectif de ce dispositif est d'améliorer la réussite scolaire parce que les étudiants sont davantage réceptifs aux attendus de l'évaluation dans une formation d'enseignement supérieur.

Dans ce partenariat, l'évaluation est alors coconstruite, c'est-à-dire que l'évaluation engendre « un processus actif en évolution utilisé par l'apprenti comme la partie d'un tout créatif où la compréhension du dispositif n'est plus un produit à accumuler » (Deeley, 2014).

Cependant, plusieurs zones d'ombres demeurent sur l'opérationnalisation de ce dispositif. Parmi les études recensées, la plupart pratiquent la co-construction de la grille d'évaluation (Falchikov, 1986 ; Stefani, 1996 ; Deeley, 2014 ; Deeley et Bovill, 2015 ; Leslie et Gorman, 2016). Dans ce

⁷ La littérature appliquée qui rentre dans le champ de la relation enseignant étudiant est moins importante que celle qui concerne l'investissement dans le capital humain. Elle s'est d'abord appliquée à étudier les facteurs de production de l'apprentissage qui ont été présentés dans la première partie de la présentation des modèles théoriques, en estimant leur effet sur la production de la réussite éducative (Rivkin et al., 2005). Ces travaux confirment les principales intuitions des modèles : les principaux déterminants de la réussite sont les préférences individuelles. L'approfondissement de cette approche a été tourné vers l'influence des notes sur le comportement d'apprentissage (Bonesrønning, 2004). Dans ce cadre, les notes sont perçues comme des formes d'incitations à l'effort d'apprentissage. Il s'agit alors de déterminer la forme optimale de la notation et du barème.

contexte, l'enseignant et les étudiants dialoguent et négocient quels sont les critères à prendre en compte pour évaluer l'effort d'apprentissage et le niveau d'acquisition des connaissances et compétences qui puissent valider l'obtention d'un module de formation. Le plus souvent, l'évaluateur propose une liste de critères, amendés par les étudiants. Puis cette grille est appliquée pour l'évaluation et peut également servir comme élément d'autoévaluation.

L'alternative à ce type de partenariat est de proposer des formules ou des menus de différentes façons d'évaluer qui sont ensuite soumises au choix des étudiants (Garside et al., 2009 ; Abdelmalak, 2016).

Enfin, mettre à contribution les étudiants est également envisageable pour créer leur propre mode d'évaluation (Doyle et al., 2017).

À la lumière de ces travaux, des marges de manœuvre semblent exister pour modifier le modèle traditionnel du contrôle des connaissances à l'université. Ces interventions ont provoqué des effets positifs sur l'engagement des étudiants (Doyle et al., 2017), leur motivation à l'apprentissage et la prise de responsabilité (Falchikov, 1986 ; Stefani, 1996 ; Deeley et Brown, 2014 ; Abdelmalak, 2016), tout en créant un environnement favorable à la compréhension de l'évaluation, son intérêt ainsi que ces objectifs. Ces expériences ont montré que le rôle de l'enseignant pouvait être modifié, en passant d'un statut autocratique à celui d'accompagnant disposé à concilier les aspects de transmission de connaissances avec celui de partenaire dans la création de l'évaluation de l'acquisition de connaissances.

Toutefois, des difficultés de compréhension des étudiants ont limité la portée des partenariats sur leur apprentissage. Ces dispositifs les sortent de leur « zone de confort » ce qui peut les déstabiliser dans leur parcours d'étude (Deeley et Bovill, 2015 ; Abdelmalak, 2016). Dans le cas où les grilles étaient co-construites avec les étudiants, il existe un risque de reproduction des critères qui prévalent dans l'évaluation, ce qui ne permet pas d'apporter d'innovations dans la façon dont est appréciée l'acquisition de connaissances.

Enfin, plusieurs aspects limitent la portée générale de ces travaux. Le premier tient aux échantillons d'étudiants considérés sur lesquels portent les interventions. Ces étudiants sont « choisis » dans des contextes plutôt favorables au changement. Par exemple, quand un module optionnel est transformé, les étudiants ont un intérêt pour l'enseignement et sont prêts à accepter des modalités d'évaluation qui diffèrent du reste de leur cursus. La taille de l'échantillon réduit quant à elle la validité externe de certaines interprétations.

À cela s'ajoute le manque d'indicateurs sur la réussite des étudiants (Garside et al., 2009 ; Leslie et Gorman, 2016). Cette dimension n'est pas majoritairement présente, alors même que les auteurs semblent vouloir innover dans les pratiques d'évaluation dans les formations d'enseignement supérieur pour renforcer la réussite de leurs groupes d'étudiants. Même si la majorité des études s'accordent sur le fait que les étudiants sont davantage motivés, la question sous-jacente non résolue est celle de savoir si le gain de motivation se traduit par un gain de performance.

Enfin, parmi ces contributions, aucune n'a conçu de protocole expérimental permettant d'évaluer des effets de causalité des dispositifs sur les étudiants⁸.

⁸ Les raisons sont multiples : les chercheurs n'avaient pas les moyens de prendre en compte cette méthode dans leur protocole de recherche, par défaut d'acculturation ou par manque d'accord des responsables administratifs.

3. HYPOTHÈSES

Bien que les études aient montré des effets positifs du partenariat dans l'évaluation, il semble qu'elles soient menées dans des cadres favorables au changement et pour des étudiants plus motivés pour accepter des modifications. Ces mêmes étudiants sont les plus susceptibles de réagir positivement à ces transformations. Il serait donc judicieux de déterminer si ces résultats positifs s'observent sur un échantillon plus diversifié et que l'on pourrait comparer à un échantillon d'individus qui n'en bénéficient pas.

Dans cet article, nous cherchons à déterminer si les étudiants qui entrent dans un partenariat d'évaluation modifient leurs stratégies d'apprentissage dans un module à l'intérieur duquel le partenariat s'effectue, mais également dans des environnements qui contiennent des formes traditionnelles d'évaluation. En d'autres termes, est-ce que l'assimilation et la co-construction des règles d'évaluation pour un module, engendrant un engagement personnel à respecter des règles décidées collectivement, permettraient de favoriser un engagement suffisant pour répondre à des règles qu'ils n'ont pas décidées, mais dont ils comprennent davantage les tenants et les aboutissants et aux attentes des enseignants dans l'ensemble des modules de formation de l'enseignement supérieur ?

Nous souhaitons ainsi tester l'effet de la négociation des règles d'évaluation dans le cadre d'une relation enseignant-étudiant sur les résultats scolaires aussi bien dans les modules où sont appliquées ces nouvelles modalités de contrôle de connaissance que dans les modules où ne se pratique pas ce dispositif. Puisque la théorie économique prédit des effets positifs sur le bien-être social et que les interventions en recherche-action des sciences de l'éducation préconisent la création de partenariats évaluatifs alors nous formulons les hypothèses suivantes :

Hypothèse 1 : la négociation des règles d'évaluation améliore les résultats scolaires des étudiants dans les modules qui intègrent ce changement grâce à une meilleure assimilation des manières d'évaluer les connaissances.

Hypothèse 2 : la négociation des règles d'évaluation améliore les résultats scolaires des étudiants de façon générale, grâce à l'adoption de nouvelles normes d'apprentissage pour répondre aux attendus de l'université sur leur acquisition de connaissances et de compétences.

4. TERRAIN

Pour tester ces hypothèses, nous avons réalisé une expérimentation de terrain au sein de deux modules de formation de la licence générale de l'Unité de Formation et de Recherche (UFR) d'Économie-Gestion de l'Université Paris-Est Marne-la-Vallée (Upem)⁹.

L'UFR d'Économie-Gestion accueille chaque année environ un quart des étudiants inscrits à l'Upem, réparti à 71,6% en licence et 28,4% en master¹⁰. La majorité de l'effectif est inscrite dans

⁹ L'Upem a été créé par le décret n°91-706 du 22 juillet 1991 dans le cadre du plan de modernisation « Université 2000 ». À l'instar des quatre autres « universités nouvelles » d'Île-de-France, ainsi que de deux universités bretonnes et lilloises, elle a été créée dans le but d'absorber l'accroissement du flux d'étudiants dans l'enseignement supérieur consécutif de la démocratisation de l'enseignement. Ce plan de construction de nouvelles universités s'est accompagné d'une prise en considération de l'aménagement du territoire afin de contribuer au développement des infrastructures économiques et techniques, des entreprises et des collectivités territoriales.

¹⁰ L'université dispose de douze composantes de formation, dont cinq instituts (électronique et informatique ; urbanisme ; ingénierie des services ; sciences appliquées ; universitaire de technologie) et sept UFR (ingénieur ;

la licence générale d'Économie-Gestion qui offre un tronc commun de formation les deux premières années et des parcours de spécialisation en économie générale et en gestion la troisième année.

Sur le plan de la réussite, les étudiants de l'UFR d'Économie-Gestion obtiennent un meilleur taux de passage en deuxième année (55%) que les étudiants de l'UFR de Sciences et techniques des activités physiques et sportives (STAPS), Mathématiques et Lettres Arts Communication et Technologie (LACT). Il est néanmoins inférieur à celui des Langues. L'UFR d'Économie-Gestion se classe également à la deuxième position du taux de validation de la licence en trois années (34%). Ces scores démontrent une marge de progression importante pour améliorer le sort des étudiants accueillis dans la licence.

Sur les raisons qui expliqueraient la difficulté des étudiants à réussir dans les formations d'Économie-Gestion, la littérature renvoie aux différents profils de bacheliers qui ne sont pas tous formés au même niveau aux outils mathématiques qui sont utiles pour la capacité d'abstraction, de modélisation et de logique, aussi bien en économie qu'en gestion (Arnold et Straten, 2012). Ainsi, à caractéristiques sociodémographiques équivalentes, un bachelier scientifique réussira mieux qu'un bachelier issu de la filière Science et technologies de management et de gestion (STMG) parce qu'il aura reçu davantage d'heures en mathématiques pour un programme plus technique. La littérature cite également les différents niveaux d'acquisition d'une culture générale qui est mobilisée à divers degrés dans la formation.

Le phénomène de l'échec à l'université est pris en considération par l'équipe dirigeante et l'équipe pédagogique qui a instauré plusieurs dispositifs successifs : une remise à niveau en mathématiques qui permet à des groupes d'étudiants de revoir ou d'approfondir les principales notions enseignées dans le programme du lycée ; un module commun à l'ensemble de la promotion se positionne sur les deux premières années de la licence et bénéficie d'un crédit d'heure au titre de l'innovation pédagogique et de l'insertion professionnelle ; un outil de gestion informatisé intitulé « Fichier de Gestion des Absences et des Résultats/notes de TD (Figaro) » réalise un suivi particulier de l'assiduité des étudiants pendant les travaux dirigés et s'intègre dans les modalités de contrôle des connaissances ; un module de méthodologie de travail universitaire.

a. L'inertie de l'évaluation de la participation

À contrario du développement de dispositifs pédagogiques innovants, les instruments d'examen des connaissances demeurent imperméables au changement. Les règles en la matière sont quasiment communes à l'ensemble de la maquette de la licence.

langues et civilisations ; lettres, arts, communications et technologie ; mathématiques ; sciences économiques et de gestion ; sciences humaines et sociales ; sciences et techniques des activités physiques et sportives).

Lors de l'année 2017-2018, 12 214 étudiants étaient inscrits dans ses formations, soit quatre fois plus que lors de sa création (Gausson, 2018). Le cycle Licence est celui qui accueille le plus d'étudiants (70% des effectifs) et il se distingue par une importante part d'inscrits en formation professionnelle (22%). Le taux de boursier (22%) est inférieur au taux national (35%), ce qui s'explique par la présence importante d'étudiants apprentis. Les étudiants proviennent majoritairement du bassin de recrutement de l'académie de Créteil (88%). Enfin, la présence d'étudiants étrangers se situe au même niveau que la proportion nationale (13%).

Le taux de réussite de l'Upem se situe au même niveau que le taux national : à titre d'exemple, 33% des néo-bacheliers inscrits en 2011-2012 validaient leur licence en trois ans et 35% des étudiants ne se réinscrivent pas au bout de la première année (Théophile, 2017). L'Observatoire des formations et des insertions professionnelles des étudiants (Ofipe) calcule également les taux de passage en deuxième année : entre 2007 et 2016, il fluctue autour de 45%.

Lorsqu'un module dispose d'une combinaison cours magistral / travaux dirigés, les étudiants sont examinés au minimum deux fois par semestre et par module au titre du contrôle continu. La plupart du temps, ces évaluations sont soumises à une règle de traitement équitable, c'est-à-dire que, dans la mesure du possible, ces évaluations sont similaires dans leurs formes et leurs contenus. Dans cette configuration, le contrôle continu représente la moitié de la note finale qui permet de valider l'acquisition d'ECTS. L'autre moitié est représentée par un contrôle terminal, la plupart du temps, sous une forme d'examens partiels effectués par tous les étudiants au même moment à la fin du semestre. Lorsqu'un module s'effectue sous la forme d'un cours magistral, l'examen partiel unique est appliqué.

La forme des examens varie peu et la primauté est donnée aux examens écrits. L'essence de l'évaluation reste identique et respecte le modèle traditionnel de l'examen dans la relation étudiant-enseignant : l'enseignant conserve l'initiative de l'évaluation et l'étudiant ne dispose pas de marges de manœuvre pour négocier une application alternative de l'évaluation. Dans le cadre des modules où le cours magistral est couplé avec des travaux dirigés, l'enseignant-référent adopte une forme d'évaluation qui est ensuite appliquée par les chargés de travaux dirigés.

Dans cette architecture globale de l'évaluation, nous avons diagnostiqué un point de tension qui provoque une insatisfaction généralisée. En effet, dans le cadre du contrôle continu il existe, dans la majorité des modules, une part attribuée à l'évaluation « générale » des étudiants qui complète le jugement effectué sur leur niveau de connaissances factuelles lors des examens semestriels. Il s'agit de l'évaluation de la participation des étudiants. Selon les choix d'évaluation dans les modules, ce score peut être compris entre 10% et 30% de la note finale.

Dans l'UFR d'Économie-Gestion, elle ne bénéficie pourtant pas d'une définition commune et partagée. Elle peut être conçue pour comptabiliser le nombre d'interventions orales des étudiants dans les modules ou le nombre de rendu volontaire d'exercices contenus dans les fascicules de travaux dirigés préparés pour chaque séance.

Aux yeux des étudiants, l'ingénierie de l'évaluation de la participation est opaque, c'est-à-dire qu'elle n'est visible qu'en comptabilisant une moyenne donnée à la fin du semestre. Elle n'influence que rarement le comportement qu'elle est censée instituer à savoir la participation active des étudiants dans les travaux dirigés. Aux yeux des enseignants chargés de travaux dirigés, c'est davantage une case à remplir qu'un outil d'évaluation des connaissances ou des compétences.

Considérer d'une manière différente cette partie de l'évaluation en contrôle continu demeure pourtant possible. Participer signifie collaborer à quelque chose, c'est-à-dire s'associer à une activité dans le but de réaliser un objectif commun. La participation est donc un concept d'action. Celle-ci repose sur des capacités individuelles particulières comme l'autonomie, la prise d'initiative, la réflexion, le jugement critique, l'analyse et la résolution de problèmes. Par conséquent, considérer la participation, c'est finalement considérer l'ensemble des compétences et dispositions que les étudiants seraient en mesure d'acquérir dans une formation d'enseignement supérieur.

De plus, la participation repose sur l'idée de la contribution qui raisonne avec la notion de coopération. La coopération peut s'effectuer avec les autres membres du groupe de travaux dirigés ou bien avec l'enseignant. Évaluer la participation, c'est finalement prendre en considération les interactions qui se jouent dans une salle de classe et que nous avons décrites dans la relation entre l'enseignant et les étudiants. Elle permet de mettre en avant l'idée que

l'étudiant est un acteur à part entière, qu'il joue un rôle dans le processus d'apprentissage et d'instruction et que sa parole prend autant de valeur que celle du tuteur.

In fine, cette notion évoque l'engagement personnel pour une tâche. Cela permet de prendre en compte le positionnement individuel de chaque étudiant dans un processus plus large.

En définitive, c'est une dimension plus vaste que la comptabilisation du nombre d'interventions. Il est possible de prendre en compte la façon dont se déploient les interactions entre étudiants, s'il existe de l'entraide ou au contraire de la rivalité. Il est possible de caractériser le fond de ces interventions : est-ce davantage prendre en considération le nombre de réponses justes à une question ? Est-ce prendre en compte le nombre d'erreurs ? Enfin dans cette réflexion, il faut être en mesure de caractériser la pondération à appliquer à chaque critère pris en compte dans chaque plan de la dimension participative. À l'inverse de considérations uniques et uniformes, la pondération et la critérisation peuvent ainsi être évolutives et dépendantes du contexte de chaque groupe de formation dans un module donné.

b. Protocole expérimental

L'expérimentation s'est déroulée lors du premier semestre de l'année universitaire 2017-2018. La séquence du traitement devait permettre de faire émerger une discussion et une négociation par étape entre les étudiants et l'enseignant sur les modalités de l'évaluation de leur participation dans les travaux dirigés.

i. Traitement expérimental

Lors de la première séance de travaux dirigés, qui est dédiée généralement à la présentation des objectifs, du contenu, des méthodes de travail et du calendrier des examens du contrôle continu, le chargé de travaux dirigés responsable de l'expérimentation explicitait les conditions dans lesquelles s'effectuerait l'évaluation de la participation. La proposition de participation à l'expérimentation était énoncée à chaque groupe de travaux dirigés lors de la première séance en ces termes : « Dans le cadre d'une étude sur l'engagement et la réussite des étudiants, votre groupe de travaux dirigés de XXX a été sélectionné pour expérimenter une nouvelle manière d'évaluer la participation des étudiants. L'objectif de cette expérience est de vous intégrer à la construction de la grille d'évaluation et de la pondération de chaque critère d'évaluation de la participation. Vous serez amenés à proposer vos propres critères d'évaluation et à amender ceux qui sont proposés par votre chargé de travaux dirigés. Une fois la grille validée, elle constituera la base de votre évaluation de la participation qui compte pour 10% de la note de contrôle continu, ce qui représente au maximum 2 points sur 20 ». Puis, l'enseignant sollicitait l'accord des étudiants pour participer ou non à cette expérimentation. Tous les étudiants acceptaient de participer au processus. À la fin de cette introduction, le chargé de travaux dirigés invitait les étudiants à réfléchir aux modalités d'évaluation de la participation sans aucune autre directive, pour qu'elles soient discutées lors de la séance suivante.

Au début de la deuxième séance de travaux dirigés, l'enseignant demandait aux étudiants de donner leurs idées sur ce qu'ils pensaient être une bonne façon de juger la participation. Le pouvoir de décision était donc d'abord dans les mains des étudiants qui avaient pu laisser libre cours à leurs réflexions, qu'elles soient saugrenues ou pas. Ce principe permettait de ne pas créer de jugements préconstruits et de ne pas restreindre la créativité ou l'imagination des groupes. Au bout de cinq minutes, la discussion était close, une liste d'éléments consensuels entre les étudiants était arrêtée et la suite de la séance se déroulait normalement.

Entre la deuxième semaine et la troisième semaine de travaux dirigés, l'enseignant disposait d'un temps de réflexion pour valider les critères débattus par les classes et amender les listes de critères. Les critères supplémentaires permettaient de considérer des caractéristiques de la participation que les étudiants n'avaient pas en tête. Dans cette phase, l'enseignant décidait également de la pondération des critères qui était commune à l'ensemble des travaux dirigés. L'idée était de considérer chaque critère à une valeur identique, qu'il soit décidé par les étudiants ou proposé par l'enseignant, c'est-à-dire que chaque critère donnait le même nombre maximal de points. Ensuite pour pouvoir juger de l'attribution des scores, une dimension quantitative était énoncée : si le critère se vérifiait durant les deux tiers des séances de travaux dirigés, le critère donnait le maximum de points (20) ; si le critère se vérifiait durant le tiers des séances de travaux dirigés, le critère donnait la moitié de points (10) ; si le critère ne se vérifiait jamais, le critère ne donnait aucun point (0). La moyenne arithmétique était réalisée pour constituer la note de participation.

Au début de la troisième séance de travaux dirigés, l'enseignant distribuait la liste des critères validés issus de la discussion des étudiants et du choix de l'enseignant ainsi que la méthode de notation des critères. Une nouvelle discussion s'enclenchait pour apporter des modifications sur la liste finale. À l'issue de cette discussion, un vote permettait de valider l'adoption de la liste. In fine, les cinq groupes de travaux dirigés tests ont obtenu une grille d'évaluation qui était spécifique à chaque discussion.

Au cours des séances suivantes, le chargé de travaux dirigés tenait les comptes des interactions qui rentraient dans la grille d'évaluation à l'aide d'un tableur. Il ne mentionnait plus particulièrement le fait de participer, laissant les étudiants libres de se conformer à la règle négociée.

À la fin du semestre, une note était calculée pour chaque étudiant et elle était transmise à l'administration pour le calcul de la note de contrôle en continu. Lors de la dernière séance qui est organisée pour rendre les seconds examens de connaissances, l'enseignant communiquait à chaque étudiant sa grille personnelle qui rappelait les principes de l'expérimentation, et qui mettait en valeur les critères validés et la note finale. Un commentaire général et/ou spécifique à chaque critère était également rédigé afin de donner des recommandations pour la suite du cursus de formation.

ii. Assignation aléatoire

Trois modules d'introduction à l'économie et deux modules de microéconomie 2 ont été choisis pour cette expérimentation. Le choix des modules a été décidé d'une part, de manière pragmatique parce qu'ils comptaient parmi le service d'enseignement d'un attaché temporaire d'enseignement et de recherche, d'autre part à partir de préoccupations sur la volonté d'améliorer les résultats des étudiants dans ces modules peu favorables à la réussite. De plus, mettre en place le dispositif expérimental dans les groupes d'un seul chargé de travaux dirigés permettait de contrôler les biais pédagogiques propres à chaque tuteur.

Les groupes ont été distribués aléatoirement au chargé de travaux dirigés par le secrétariat pédagogique. Cette répartition est déterminée par l'emploi du temps personnel de l'enseignant et par l'organisation générale du planning de formation qui dépend des salles disponibles aux créneaux choisis et du chevauchement d'autres modules aux horaires considérés. Nous sommes donc assurés qu'il n'y a pas eu de sélection particulière dans l'attribution des travaux dirigés selon la liste des étudiants.

De plus, les étudiants sont répartis dans les groupes selon un panachage aléatoire réalisé par le secrétariat pédagogique. Cette répartition dépend à la fois du choix des unités d'enseignements libres (les options) de manière à rassembler au sein d'un même groupe de travaux dirigés des étudiants suivant la même option et de l'ordre alphabétique de leur nom de famille. Ainsi, les étudiants ne « choisissent » pas leur emploi du temps. Enfin, après l'initialisation des groupes, aucun changement n'est autorisé par les responsables de formation.

Le groupe témoin est constitué par les cinq autres groupes de travail dirigés de première année et les quatre groupes de travail dirigés de deuxième année.

En définitive, on se rapproche d'une expérimentation contrôlée grâce à l'affectation aléatoire dans des groupes répartis aléatoirement dans le planning de formation. Dans le cadre de l'expérimentation du Projet Voltaire (Bellity et al., 2017), ce même processus d'affectation des étudiants aux groupes de travail dirigés a été assimilé à une assignation aléatoire.

5. DONNÉES

Nous souhaitons évaluer l'effet de ce traitement sur les résultats des étudiants à tous les examens du premier semestre, que ce soit en contrôle terminal ou en contrôle continu, à la fois dans les modules dans lesquels a été conduite l'expérimentation et pour les modules sans modification.

a. Collecte

Pour réaliser cette évaluation, nous utilisons une extraction des données issues de l'application pour l'organisation et la gestion des enseignements et des étudiants (Apogée) pour les inscrits au premier semestre 2017-2018 qui contient les noms et prénoms des étudiants, leur date de naissance, et pour chaque module, les notes finales du contrôle continu et du contrôle terminal¹¹. Si les étudiants ne respectaient pas les modalités de contrôle des connaissances, les mentions de défaillance étaient renseignées.

Pour les étudiants de première année, nous utilisons leurs résultats du premier semestre de la première session. Il n'existait pas de mention spécifique pour identifier les étudiants qui étaient redoublants dans la promotion. Nous avons donc utilisé les résultats de la promotion 2016-2017 de première année pour identifier les redoublants.

Pour les étudiants de deuxième année, nous procédons de la même façon, à savoir l'édition d'une extraction qui contient les résultats de la première session du premier semestre de l'année 2017-2018, ainsi qu'une extraction des résultats du premier semestre de l'année 2016-2017 pour identifier les redoublants.

Comme nous disposons d'informations de la première année de la promotion 2016-2017, nous pouvons identifier les étudiants qui sont passés de la L1 à la L2 ainsi que les personnes qui proviennent d'une autre formation en réorientation. Pour les premiers, nous disposons ainsi

¹¹ L'application est structurée en « unités de gestion » des informations issues des inscriptions administratives (état civil, frais d'inscription, sécurité sociale) et pédagogiques (rattachement à un diplôme, une année, à un ensemble de modules de formation), des modalités de contrôle des connaissances et des épreuves (barème, nombre de sessions, règles de calcul, planning), des résultats et des notes (saisie des notes et de la situation d'ajournement ou de défaillance) et de la production des diplômes (procès-verbaux, annexe au diplôme, etc. À cause du nombre de renseignements sensibles et pour éviter des fraudes aux diplômes liées à l'utilisation malhonnête de l'application, les accès à Apogée sont restreints à un faible nombre d'utilisateurs. Nous avons fait appel au secrétariat pédagogique de la licence générale pour avoir accès à ces informations.

d'informations sur leurs résultats scolaires antérieurs, tandis que pour les seconds, ces informations ne sont pas disponibles.

Nous utilisons également les bases de gestion de l'outil Figaro. Ce dernier permet d'identifier les groupes de travail dirigés dans lesquels les étudiants sont inscrits, leur présence hebdomadaire et la décomposition des notes du contrôle continu, ainsi que leur pondération pour chaque module.

Enfin, nous avons collecté les informations issues de la plateforme Admission Post-Bac (APB) remplacée au printemps 2018 par la plateforme ParcoursSup. APB était conçue pour la saisie des vœux d'inscription des lycéens futurs bacheliers ou des personnes en reprise d'études dans les formations d'enseignement supérieur. La personne responsable de la gestion des candidats a utilisé cette plateforme pour extraire des informations sur la filière d'origine du baccalauréat des étudiants et sur leurs notes obtenues au baccalauréat. Cependant, nous disposons de ces informations uniquement pour les étudiants dont les vœux ont été validés à la fin du processus de sélection. Par conséquent, il n'est pas possible d'obtenir ces informations pour les étudiants qui se sont réorientés grâce à des dispositifs internes à l'université, ou des bacheliers ayant été affectés par le rectorat pendant la phase complémentaire d'assignation. Ces caractéristiques sociodémographiques, au même titre que le lieu de domicile, ne font pas l'objet d'une saisie systématique du secrétariat pédagogique.

Comme l'expérimentation de terrain est introduite dans un contexte réel et pas en laboratoire, elle peut bouleverser l'organisation de la formation. Le tirage au sort pour l'assignation au dispositif expérimental pose notamment une question d'ordre morale. En effet, l'attribution aléatoire de la condition expérimentale produit une rupture d'égalité « des chances » ex post en privant les étudiants de ressources temporaires qui pourraient améliorer (détériorer) leur situation. En l'occurrence dans le cadre de la licence, c'est prendre le risque d'affaiblir (d'augmenter) les chances de réussite aux examens et in fine de voir les étudiants être pénalisés (favorisés) dans l'obtention de leur diplôme. Cette rupture du principe d'égalité justifie le caractère temporaire d'une expérimentation (L'Horty et Petit, 2011).

C'est pourquoi afin de se prémunir contre d'éventuels biais d'organisation, l'expérimentation a bénéficié de la constitution d'un comité de suivi qui était composé de la responsable du cycle de la licence, du responsable de la première année et de la responsable de la deuxième année. Ce comité a permis d'analyser le dispositif et de valider chaque étape du processus pour s'assurer de la pertinence éthique. Les enseignants-référents des modules étaient également tenus au courant du dispositif.

À partir des recommandations de ce comité, de la prise en compte du règlement général de la protection des données (RGPD) qui renforce les prérogatives des individus collectant des informations personnelles, ainsi que des préconisations en matière de démarche scientifique du Centre national pour la recherche scientifique (CNRS)¹², nous avons demandé le consentement

¹² Notamment le guide « [Pratiquer une recherche intègre et responsable](#) » du Comité d'éthique du CNRS (Comets) qui préconise la protection des données à caractère personnel à travers cinq principales recommandations : la sécurisation des fichiers ; la confidentialité des données ; le principe de finalité pour indiquer le but de la collecte ; le principe de proportionnalité qui permet de fixer la quantité et la durée de la collecte ; l'information des sujets d'étude.

¹² À l'instar des quatre autres « universités nouvelles » d'Île-de-France, ainsi que de deux universités bretonnes et lilloises, elle a été créée dans le but d'absorber l'accroissement du flux d'étudiants dans l'enseignement supérieur consécutif de la démocratisation de l'enseignement. Ce plan de construction de nouvelles universités s'est accompagné d'une prise en considération de l'aménagement du territoire afin de contribuer au développement des infrastructures économiques et techniques, des entreprises et des collectivités territoriales.

des étudiants quant à l'utilisation de leurs informations personnelles dans le cadre du dispositif expérimental.

Ainsi, les promotions de première année et de deuxième année ont signé un accord de consentement de l'utilisation de leurs informations dans lequel était mentionné une description du cadre du projet de recherche et de la finalité de la collecte, que les données collectées portaient sur les résultats scolaires ainsi que sur d'autres informations administratives, et qu'il était possible de consulter ou de supprimer l'accès à ces informations.

Dans le cadre de cette démarche, nous avons pu recueillir le consentement de 70% des étudiants des deux promotions. En revanche, une des limites de cette décision est que plusieurs étudiants des groupes traités ont refusé l'accès à leurs informations personnelles. En respectant leur volonté, nous n'avons donc pas utilisé leurs résultats dans le cadre de notre étude. Cela peut donc constituer un biais dans nos interprétations finales.

b. Échantillon

Lors du premier semestre de l'année 2017-2018, la promotion de première année de licence générale d'Économie-Gestion comptait 289 inscrits et la promotion de deuxième année comptait 202 inscrits.

Parmi ces inscrits, nous enlevons les étudiants qui relèvent de la situation « prise en compte pour calcul » (Prc) : ce sont des étudiants qui sont inscrits pédagogiquement pour l'année en cours, mais qui ont validé l'un des deux semestres l'an passé. Dans ce cas, ils redoublent le semestre qu'ils n'ont pas validé et peuvent bénéficier d'un passage dans l'année supérieure. Il existe donc des étudiants qui ne sont pas présents dans les travaux dirigés du premier semestre. Par conséquent, nous les retirons de l'analyse pour ne conserver que les étudiants qui sont en travaux dirigés puisque le traitement s'applique dans ce contexte.

De plus, comme nous l'avons expliqué dans la partie de présentation du protocole expérimental, nous retirons également de l'analyse les étudiants qui n'ont pas consenti à donner un accès à leurs informations personnelles. Les caractéristiques des étudiants du groupe traité et du groupe témoin sont renseignées dans le tableau 1 et le tableau 2.

Tableau 1 : Statistiques descriptives et test d'équivalence entre le groupe traité et le groupe témoin des étudiants de première année de licence d'Économie-Gestion 2017-2018

	Groupe traité	Groupe témoin	Différence	Test de comparaison	
				Test de Wald	Test de Student
Part de femmes	45,455 (5,674)	45,283 (4,835)	-0,172 (7,454)	-0,023	
Part de redoublants	11,688 (3,661)	10,377 (2,962)	-1,311 (4,672)	-0,2806	
Part de réorientation	19,481 (4,513)	19,811 (3,871)	0,331 (6,188)	0,0556	
Part de néo-entrants	68,831 (5,278)	69,811 (4,459)	0,98 (6,900)	0,142	
Taux de BAC ES	62,338 (5,522)	63,208 (4,684)	0,87 (7,236)	0,1202	
Taux de BAC S.	16,883 (4,269)	16,981 (3,647)	0,098 (5,617)	0,0175	
Taux de BAC autre (L, STMG, P)	11,688 (3,661)	11,321 (3,077)	-0,368 (4,772)	-0,077	
Âge (au début du traitement)	19,221 (0,996)	19,120 (1,158)	-0,101 (0,163)		-0,6165
Moyenne générale au baccalauréat	11,949 (1,438)	12,014 (1,591)	-0,065 (0,280)		0,2329
Nombre de groupes de travail dirigés	3	5			
Nombre d'étudiants	77	106			

Note : Les données entre parenthèses sont les erreurs-types pour les variables discrètes et les écarts-types pour les variables continues.

Il y a davantage d'étudiants inscrits dans la formation que d'étudiantes. La part de redoublants du premier semestre concerne un étudiant sur dix. Deux étudiants sur dix n'étaient pas en terminale l'année précédente. La majorité de la promotion est diplômée d'un baccalauréat économique et social. L'âge moyen au début de l'expérimentation est de 19 ans, ce qui correspond à un an de retard avec l'âge « normal » des étudiants à l'entrée de la licence. Cela reflète bien l'idée que la promotion contient des étudiants redoublants ou en réorientation. La moyenne générale obtenue au baccalauréat est de 12/20, ce qui équivaut à la mention « assez bien », et la promotion ne semble pas spécialement hétérogène sur son niveau antérieur.

Nous notons que la répartition correspond à une allocation de 40% des étudiants dans le groupe traité et 60% des étudiants dans le groupe témoin. Les tests pratiqués sur les caractéristiques observables dans les deux groupes nous permettent de ne pas rejeter l'hypothèse de nullité des écarts sur les caractéristiques observables au seuil de 5%.

Tableau 2 : Statistiques descriptives et test d'équivalence entre le groupe traité et le groupe témoin des étudiants de deuxième année de licence d'Économie-Gestion 2017-2018

	Groupe traité	Groupe témoin	Différence	Test de comparaison	
				Test de Wald	Test de Student
Part de femmes	45,565 (7,344)	41,111 (5,187)	-4,454 (8,990)	-0,5066	
Part de redoublants	4,348 (3,007)	17,778 (4,030)	13,43** (5,028)	-2,187	
Part de réorientation	6,522 (3,640)	10,000 (3,162)	3,478 (4,822)	0,6766	
Part de néo-entrants	89,130 (4,589)	72,222 (4,721)	-16,908** (6,582)	-2,2499	
Taux de BAC ES	34,783 (7,022)	34,444 (0,501)	-0,338 (8,619)	-0,0392	
Taux de BAC S.	21,739 (6,082)	18,889 (4,126)	-2,85 (0,735)	-0,3942	
Taux de BAC autre (L, STMG, P)	4,348 (3,007)	2,222 (1,554)	-2,126 (3,385)	-0,6941	
Âge	20,140 (0,953)	20,084 (0,910)	-0,055 (0,168)		-0,3294
Moyenne générale au Baccalauréat	12,580 (1,582)	12,281 (1,314)	-0,299 (0,350)		-1,1368
Nombre de groupes de travail dirigés	2	4			
Nombre d'étudiants	46	90			

Note : Les données entre parenthèses sont les erreurs-types pour les variables discrètes et les écarts-types pour les variables continues. ** : p-value < 0,05.

En ce qui concerne la promotion de deuxième année, nous observons qu'il y a également davantage d'étudiants inscrits que d'étudiantes. La part de redoublants atteint 13%, tandis que le taux d'étudiants en réorientation est de 8%. Lorsque la filière du baccalauréat est renseignée, le mode est la filière économique et sociale. L'âge moyen est de 20 ans, ce qui correspond de nouveau au phénomène de parcours scolaire avec retard. Lorsque la moyenne du baccalauréat était renseignée, les étudiants décrochaient une note générale de 12/20.

Nous notons que la répartition correspond à une allocation de 30% de la promotion dans le groupe traité et 70% dans le groupe témoin. À contrario de la promotion de première année, le groupe témoin est différent du groupe traité sur la répartition des parcours antérieurs des étudiants. Il y a significativement moins de redoublants dans le groupe traité. C'est donc une caractéristique qu'il faudra tenir compte dans les analyses suivantes.

6. MÉTHODE D'IDENTIFICATION DES EFFETS

a. Les indicateurs d'intérêt

Notre variable d'intérêt concerne la réussite académique. Nous approchons cette dimension à l'aide de deux indicateurs.

Le premier indicateur considère les résultats semestriels du module dans lequel est intervenu le processus de négociation des règles d'évaluation de la participation aux travaux dirigés. Trois notes sont retenues dans le cadre de l'expérimentation : le premier examen en contrôle continu ; le second examen en contrôle continu ; l'examen partiel en contrôle terminal. Nous nous

restreignons ainsi à l'effet de la condition expérimentale sur la réussite dans le module, à diverses étapes de l'effort d'apprentissage des étudiants.

Le second indicateur est une mesure plus globale de la réussite qui prend en compte le fait d'avoir validé ou non le semestre lors de la première session d'examen. Il s'agit dans ce cas d'évaluer l'impact de la condition expérimentale sur la stratégie générale des étudiants et de déterminer si ce dispositif ponctuel pour un seul module a permis de modifier les dispositions des étudiants pour leur stratégie de réussite à l'université.

b. L'effet moyen du traitement

L'évaluation causale des effets est permise grâce au modèle de Rubin (1974). Dans ce cadre, nous cherchons à estimer l'effet moyen du traitement, la négociation de la règle d'évaluation du partenariat, sur les notes obtenues lors du contrôle continu et lors du contrôle terminal.

Dans un premier temps, nous comparons les écarts de notes moyennes obtenues aux trois examens du module par le groupe traité et le groupe témoin. Nous réalisons un test d'hypothèse de nullité des écarts observés en utilisant la statistique de Student.

Dans un second temps, nous utilisons les résultats obtenus dans les autres modules afin de contrôler de l'hétérogénéité de la difficulté des autres composantes de la formation. Dans ce contexte, on peut alors se rapprocher d'un modèle d'estimation en double différence :

$$Y_{im} = \alpha_1 \cdot G_i + \alpha_2 \cdot T_m + \alpha_3 \cdot (G_i \cdot T_m) + \beta + \epsilon_{im}$$

où Y_{im} correspond à la note d'un étudiant i à un examen (de contrôle continu ou terminal) d'un module m . La variable G_i est une indicatrice d'appartenance au groupe traité ($G_i = 1$) ou au groupe témoin ($G_i = 0$). La variable T_m correspond à l'indicatrice du module traité. La variable d'interaction $G_i \cdot T_m$ correspond à l'appartenance au groupe traité dans le module traité. Par conséquent, α_3 mesure l'indicateur de double différence suivant :

$$\alpha_3 = (\overline{Y_{11}} - \overline{Y_{01}}) - (\overline{Y_{10}} - \overline{Y_{00}})$$

où la première différence correspond à l'écart de notes observé uniquement pour le module traité et la seconde différence à l'écart de notes observé pour les autres modules. De cette manière, nous tenons compte du résultat des étudiants dans les autres modules et cela nous permet de capturer un effet pur du traitement dans le module. Si $\alpha_3 > 0$, cela indique que les étudiants du groupe traité ont obtenus de meilleurs résultats dans le module traité par rapport au groupe témoin et leurs résultats dans les autres modules. Nous pourrions ainsi attribuer à l'environnement de négociation l'effet causal de cet effet.

Les autres paramètres du modèle mesurent d'autres indicateurs d'intérêt :

$$\beta = \overline{Y_{00}}$$

qui correspond à la note moyenne obtenue par le groupe témoin dans les modules non traités ;

$$\alpha_1 = \overline{Y_{10}} - \overline{Y_{00}}$$

qui identifie l'écart de notes moyennes entre le groupe traité et le groupe témoin pour les modules de formation non traités ;

$$\alpha_2 = \overline{Y_{01}} - \overline{Y_{00}}$$

qui mesure la différence de notes moyennes obtenues pour le groupe témoin entre le module traité et les modules non traités.

Nous estimons ce modèle par la méthode des moindres carrés ordinaires, en tenant compte d'une éventuelle hétéroscédasticité.

c. Modèle dichotomique pour mesurer l'effet sur la validation du diplôme

Pour estimer l'effet du traitement sur le fait de valider le semestre, nous recourons à un modèle dichotomique de type logit. Il s'écrit de la manière suivante :

$$\text{logit } P(Y_i = 1|X_i) = \alpha_1 \cdot G_i + \beta_1 \cdot X_i + \alpha_0$$

où Y_i est l'indicatrice de validation du semestre qui vaut $Y_i = 1$ lorsque l'étudiant est admis, et $Y_i = 0$ s'il ne l'est pas. Nous ne rentrons pas dans des considérations de modélisation polytomiques, qui permettraient de tenir compte des résultats en termes de défaillance, d'absences justifiées ou d'étudiants ajournés.

La validation du semestre dépend des caractéristiques observables des étudiants qui rentrent dans la variable X_i : le type de baccalauréat, le genre, le fait d'avoir redoublé. La variable G_i permet d'identifier l'appartenance au groupe témoin. Dans ce cadre, α_1 estime l'effet du traitement sur la réussite. Nous estimons ce modèle grâce à la méthode du maximum de vraisemblance.

7. RÉSULTATS

Dans cette partie nous présentons les résultats de l'expérimentation sur les indicateurs de réussite académique. Nous illustrons une forme de grille d'évaluation de la participation à l'issue de la négociation en annexe (voir annexe b) et nous réalisons un test d'équivalence sur les notes obtenues sur la participation par le groupe traité (voir annexe c).

Nous effectuons les évaluations d'impact en distinguant les deux promotions, de première année et de deuxième année, afin de tenir compte de la spécificité de chaque période de formation. De plus, nous tenons compte des résultats à l'issue de la première session d'examen, ce qui implique que les résultats finaux de la deuxième session ne sont pas considérés dans ces analyses. La session de rattrapages se réalise en effet à la fin de l'année étudiante, après les examens du second semestre. Par conséquent, les étudiants réalisent des arbitrages différents lors de cette période, pour maximiser les chances de valider une unité d'enseignements. Il serait possible de mesurer l'effet de durabilité de l'expérimentation sur les étudiants au rattrapage, mais le faible nombre d'observations réduirait le pouvoir statistique de mesure.

Dans un premier temps, nous estimons les effets de l'expérimentation sur les résultats aux trois examens de connaissances. D'une part sur le module traité, sans tenir compte des résultats des étudiants dans les autres matières, ce qui permet de déterminer l'effet moyen du traitement, et d'autre part en tenant compte des résultats dans les autres matières, ce qui nous permet de contrôler les stratégies de réussite des étudiants dans l'ensemble de la formation. Dans un second temps, nous évaluons l'impact de l'expérimentation sur les chances de validation du premier semestre pour les étudiants de première année et du troisième semestre pour les étudiants de deuxième année.

a. La réussite aux examens de contrôle continu et terminal

i. Les résultats d'examens des étudiants de première année

Les résultats de l'expérimentation sur la réussite aux examens des étudiants de première année sont inscrits dans le tableau 3¹⁵.

Tableau 3 : Évaluation d'impact de la condition expérimentale sur les résultats des étudiants de première année

	Résultat CC1	Résultat CC1	Résultat CC2	Résultat CC2	Résultat CT	Résultat CT
Différence simple traité – témoin sur le module traité	1,043*** (0,389)		1,787*** (0,439)		0,468 (0,421)	
Différence simple traité – témoin sur les modules non traités		-0,469 (0,339)		0,814** (0,335)		0,603** (0,282)
Différence pour le groupe témoin entre le module traité et ceux non traités		- 0,945*** (0,344)		0,165 (0,368)		- 4,608*** (0,350)
Double différence		1,512*** (0,516)		0,973* (0,551)		-0,134 (0,505)
Référence : Groupe témoin	10,08*** (0,273)	11,03*** (0,209)	12,70*** (0,297)	12,54*** (0,219)	6,731*** (0,296)	11,34*** (0,189)
Nombre d'observations	174	846	171	831	166	1138

Note : les erreurs-types sont entre parenthèses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Nous constatons un effet positif et significatif de la condition expérimentale, c'est-à-dire le fait d'avoir participé à un processus de co-construction de l'évaluation de la participation dans le module d'introduction à l'économie, sur les résultats du module, à la fois lors du premier examen en contrôle continu et lors du deuxième examen en contrôle continu (colonne 1 et colonne 3). Nous observons même une amélioration des résultats entre les deux examens, plus importante pour le groupe traité, mais cette progression des résultats n'est pas significative.

Lorsque l'on tient compte de la performance des étudiants dans les examens des matières qui n'ont pas disposé de la négociation de la règle d'évaluation de la participation, l'effet positif se maintient (colonne 2 et 4). Cependant, l'écart de réussite consécutif au traitement dans le module traité s'atténue entre les deux périodes d'examen. Nous mesurons par ailleurs que les étudiants du groupe de traitement ont été significativement meilleurs dans les modules non traités que le groupe témoin au cours du deuxième examen en contrôle continu. Enfin, les étudiants traités par le dispositif expérimental ont mieux réussi dans le module traité que dans les autres modules. Cet

¹⁵ En ce qui concerne les étudiants de la promotion de première année de licence d'Économie-Gestion de l'année 2017-2018, le premier examen en contrôle continu du module d'introduction à l'analyse économique était constitué d'un QCM qui portait sur l'acquisition de connaissances générales en économie à partir d'ouvrages de la collection *Repères* lus et étudiés chaque semaine par les étudiants. Le deuxième examen était formé de la même manière et portait sur un ensemble plus important de lectures. Le contrôle terminal contenait un QCM ainsi qu'un exercice de résumé de texte. Cette méthode de travail avait été apprise en travaux dirigés.

Dans les examens en contrôle continu, nous tenons compte des matières de mathématiques, d'anglais, d'introduction à la gestion et de méthodologie de travail universitaire. Pour toutes ces matières, les examens étaient des tests écrits. Pour les examens en contrôle terminal, nous ajoutons les résultats dans les modules optionnels (langue vivante 2, introduction à la sociologie, histoire de la pensée économique) et dans le module de construction de projet d'étude.

écart de réussite est deux fois plus important lors du second examen que lors du premier examen en contrôle continu, et il est significatif. Cependant, la progression de l'écart de réussite entre les deux périodes d'examen et entre les deux types de modules pour le groupe traité n'est pas significative.

Par conséquent, l'application d'un environnement favorable à la négociation de la règle de l'évaluation de la participation dans les travaux dirigés a permis aux étudiants de s'impliquer davantage dans le module ce qui s'est traduit par de meilleurs résultats que pour les étudiants qui n'en bénéficiaient pas.

Néanmoins, ce constat positif disparaît lorsque nous considérons les résultats lors de l'examen partiel en contrôle terminal (colonne 5 et 6). Alors que le contrôle terminal se soldait par une note moyenne de 5 points inférieurs à la note de contrôle continu, il ne semble pas que le traitement ait contribué à significativement atténuer ces résultats en baisse pour les étudiants participant à l'expérimentation. Cette absence d'effet s'observe également lorsque nous tenons compte des résultats dans les autres modules de formation. Enfin, nous observons que le groupe traité a acquis de meilleurs résultats en moyenne au contrôle terminal pour les modules où le traitement était absent.

ii. Les résultats d'examens des étudiants de deuxième année

Les résultats sur la réussite aux examens des étudiants de deuxième année sont inscrits dans le tableau 4¹⁴.

¹⁴ En ce qui concerne les étudiants de la promotion de deuxième année de licence d'Économie-Gestion de l'année 2017-2018, le premier examen en contrôle continu du module de microéconomie niveau 2 était constitué de deux exercices de modélisation de la théorie du consommateur et du producteur et de questions de mesure des connaissances acquises à partir du cours magistral. Le deuxième examen était formé de la même manière et portait sur des exercices de modélisation de la théorie de l'équilibre général. Le contrôle terminal contenait un QCM ainsi qu'un exercice de modélisation de la concurrence parfaite et imparfaite.

Dans les examens en contrôle continu, nous tenons compte des matières de macroéconomie 2, de mathématiques, d'anglais, de comptabilité et de gestion des ressources humaines. Pour toutes ces matières, les examens étaient aussi des tests écrits. Pour les examens en contrôle terminal, nous ajoutons les résultats dans les modules optionnels (langue vivante 2, sociologie du consommateur, problèmes économiques contemporains) et dans le module de construction de projet professionnel, et de droit des contrats.

Tableau 4 : Évaluation d'impact de la condition expérimentale sur les résultats des étudiants de deuxième année

	Résultat CC1	Résultat CC1	Résultat CC2	Résultat CC2	Résultat CT	Résultat CT
Différence simple traité – témoin sur le module traité	-1,137 (0,687)		- 1,845*** (0,639)		1,148* (0,634)	
Différence simple traité – témoin sur les modules non traités		0,527 (0,338)		0,0703 (0,294)		1,274*** (0,247)
Différence pour le groupe témoin entre le module traité et ceux non traités		-2,117*** (0,506)		- 4,123*** (0,471)		- 1,920*** (0,426)
Double différence		-1,665** (0,762)		-1,915*** (0,701)		-0,126 (0,677)
Référence : Groupe témoin	9,982*** (0,465)	12,10*** (0,204)	8,789*** (0,443)	12,91*** (0,167)	7,963*** (0,403)	9,884*** (0,145)
Nombre d'observations	128	771	128	771	127	1263

Note : les erreurs-types sont entre parenthèses. * $p < 0,10$, ** $p < 0,05$, *** $p < 0,01$.

L'évaluation de l'impact de la condition expérimentale sur les scores des étudiants de deuxième année conduit à des résultats opposés à ceux de première année. En effet, les étudiants du groupe traité ont moins bien réussi que le groupe témoin aux deux examens en contrôle continu, et ce résultat négatif est significatif lors du deuxième contrôle des connaissances (colonne 3). De plus, alors que les résultats diminuaient pour les deux groupes entre le premier et le deuxième examen en contrôle continu, la baisse des résultats était plus importante pour le groupe traité. En revanche, cet écart n'est pas significatif.

L'écart négatif des résultats entre le groupe traité et le groupe témoin dans le module qui bénéficiait d'un environnement plus favorable à la négociation de l'évaluation de la participation s'observe de nouveau lorsque nous tenons compte des résultats des étudiants dans les autres matières (colonne 2 et 4). En effet, alors que les deux groupes présentent des résultats similaires dans les modules de travaux dirigés sans intervention expérimentale, les étudiants traités ont obtenu significativement de moins bons résultats en microéconomie 2, au cours des deux examens en contrôle continu. Cet effet s'accroît entre les deux contrôles de connaissances. Nous notons également que les étudiants obtiennent un écart de -4 points en microéconomie par rapport aux autres modules.

Par conséquent, l'adoption d'une pratique de négociation de l'évaluation de la participation n'a pas permis d'améliorer les résultats des étudiants. Au contraire, cela a davantage renforcé leur échec en diminuant leur réussite aux examens.

En revanche, l'effet de la condition expérimentale sur le résultat des étudiants pendant le contrôle terminal est positif et significatif. Les étudiants du groupe traité ont ainsi obtenu en moyenne 1,2 point supplémentaire au partiel de microéconomie, ce qui a permis de compenser en partie les résultats obtenus dans le cadre du contrôle continu. Cependant, cet écart positif ne s'observe pas quand nous tenons compte des résultats dans les autres matières. Les étudiants sous la condition expérimentale ont ainsi mieux réussi dans les modules non traités que le groupe témoin.

L'hétérogénéité de l'impact du traitement sur la réussite des étudiants nous interpelle quant à la question de savoir s'il est possible de généraliser le dispositif. Nous y reviendrons dans la partie d'interprétation de cette évaluation d'impact.

b. La réussite en termes de validation du semestre

Dans cette section, nous estimons l'impact de la négociation de l'évaluation de la participation en travaux dirigés sur un indicateur de réussite plus global. En effet, nous souhaitons déterminer si les étudiants qui ont bénéficié de ce dispositif ont conduit des stratégies de réussite qui leur ont permis de valider leur semestre lors de la première session d'examen. Dans ce cadre, les notes seules ne jouent pas sur l'admission, la pondération des modules et les règles de compensation rentrent aussi en considération dans l'effort d'apprentissage. Ainsi, nous évaluons la capacité des étudiants à mieux tenir compte de cet environnement grâce au dispositif qui leur permettait d'acquérir des dispositions d'autonomie et de prise de décision.

i. La validation du premier semestre des étudiants de première année

Les résultats de l'estimation du modèle logit pour la réussite des étudiants de première année sont inscrits dans le tableau 5.

Tableau 5 : Estimation de l'effet du traitement sur la probabilité de valider le premier semestre pour les étudiants de première année

	(1)	(2)	(3)	(4)	(5)
Condition expérimentale	0,080 (0,073)	0,088 (0,072)	0,086 (0,070)	0,086 (0,070)	0,085 (0,067)
Âge		-0,079** (0,039)	-0,065 (0,041)	-0,051 (0,063)	-0,044 (0,058)
Sexe			0,223*** (0,070)	0,225*** (0,070)	0,194*** (0,069)
<i>Référence : Néo-entrants</i>					
Redoublants				-0,056 (0,165)	-0,023 (0,156)
Réorientation				-0,026 (0,112)	-0,053 (0,132)
<i>Référence : Baccalauréat E.S</i>					
Baccalauréat S					0,247** (0,098)
Autre baccalauréat					-0,300** (0,118)
Baccalauréat non communiqué					0,080 (0,135)
Constante	0,227 (0,196)	6,733** (3,374)	5,452 (3,670)	4,306 (5,424)	4,061 (5,393)
Pseudo R ²	0.005	0.028	0.067	0.068	0.137
<i>Critère AIC</i>	250,5	246,8	239,1	243,0	231,8
Nombre d'observations	183	183	183	183	183

Note : les effets moyens marginaux sont inscrits pour chaque variable. Les erreurs-types sont entre parenthèses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Nous constatons un effet positif, mais non significatif de la condition expérimentale sur la probabilité de valider le semestre. Cet effet se maintient avec l'inclusion de variable de contrôle. Par conséquent, nous concluons que les étudiants n'auraient pas établi de stratégie d'optimisation des règles de validation lors du premier semestre de la première année de licence lorsqu'ils ont bénéficié d'un environnement qui leur aurait permis de mieux comprendre l'ingénierie de l'évaluation.

Parmi ces variables, nous notons que les femmes ont davantage de chance de valider leur premier semestre que les hommes, de 20 points de pourcentage de différence. Les bacheliers de la filière scientifique montrent également plus de chances de valider leur premier semestre que les bacheliers de la filière économique et sociale.

ii. La validation du troisième semestre des étudiants de deuxième année

L'évaluation de l'impact de la condition expérimentale sur la probabilité de valider le troisième semestre de la licence pour les étudiants en deuxième année est inscrite dans le tableau 6.

Tableau 6 : Estimation de l'effet du traitement sur la probabilité de valider le troisième semestre pour les étudiants de deuxième année

	(1)	(2)	(3)	(4)	(5)
Condition expérimentale	0,140* (0,085)	0,148** (0,074)	0,140* (0,072)	0,148** (0,075)	0,162** (0,074)
Âge		-0,164*** (0,042)	-0,162*** (0,041)	-0,170*** (0,049)	-0,150*** (0,053)
Sexe			0,194*** (0,075)	0,196*** (0,074)	0,204*** (0,075)
<i>Référence : Néo-entrants</i>					
Redoublants				0,047 (0,116)	0,082 (0,117)
Réorientation				0,023 (0,127)	0,079 (0,132)
<i>Référence : Baccalauréat E.S</i>					
Baccalauréat S					-0,182* (0,105)
Autre baccalauréat					-0,319 (0,214)
Baccalauréat non communiqué					-0,165 (0,103)
Constante	0,314 (0,214)	16,51*** (5,152)	16,68*** (5,415)	17,42*** (6,106)	16,50** (6,571)
Pseudo R ²	0,014	0,100	0,135	0,136	0,163
<i>Critère AIC</i>	181,4	167,9	163,7	167,5	168,6
Nombre d'observations	136	136	136	136	136

Note : les effets moyens marginaux sont inscrits pour chaque variable. Les erreurs-types sont entre parenthèses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Nous observons que pour les étudiants qui sont entrés dans un partenariat évaluatif dans le module de microéconomie, leur chance de valider leur troisième semestre augmente de 14 points de pourcentage. Cet effet positif se maintient avec l'inclusion de variables de contrôle. Par conséquent, bien que leurs notes en microéconomie aient été significativement inférieures à celles du groupe témoin, les étudiants ont adopté des stratégies d'optimisation de leurs résultats dans les autres matières. Il est possible d'envisager que les étudiants aient intégré des dispositions comportementales pour prendre en compte des règles d'évaluation pour maximiser leurs chances d'obtenir de meilleurs résultats dans les modules sans environnement de collaboration sur l'implantation de l'évaluation de la participation.

Pour la promotion de deuxième année, on constate que plus les étudiants sont âgés, moins ils possèdent de chance de valider leur semestre. Le fait d'être une femme augmente la probabilité de valider le semestre de 20 points de pourcentage. De manière surprenante, les étudiants passés

par la filière scientifique du baccalauréat ont moins de chance de valider leur semestre lors de la première session. Néanmoins, l'interprétation de ce coefficient est à prendre avec précaution, parce que l'information sur le diplôme du baccalauréat était mal renseignée pour cette promotion.

8. INTERPRÉTATIONS ET DISCUSSION DES RÉSULTATS

À l'issue de notre expérimentation de la négociation des règles d'évaluation de la participation dans les travaux dirigés d'introduction à l'économie en première année et de microéconomie en deuxième année de licence d'Économie-Gestion, nous obtenons des résultats hétérogènes et des effets qui semblent opposés : un gain de performance pour les uns ; une détérioration de performance pour les autres. Si nous devons synthétiser nos résultats en une formule simple, celle-ci serait que les effets de la mise en place d'une négociation des règles d'évaluation entre les étudiants et les enseignants pour promouvoir la réussite dépendent du contexte dans lequel elle se réalise. Ce contexte se caractérise par plusieurs dimensions : la temporalité de la relation, la prise en compte des règles de fonctionnement de l'organisation, le sens de convergence des intérêts individuels et les termes de l'exécution du contrat négocié. Les conditions de ce contexte influencent le comportement des étudiants dans leurs stratégies de réussite à l'université. Cette influence se traduit par des résultats différenciés.

a. En première année, une convergence des intérêts est compatible avec la négociation de l'évaluation

Si l'on se réfère aux arguments théoriques, le partenariat et la coopération peuvent aboutir à une situation socialement optimale si les deux parties de la relation d'instruction considèrent que l'enseignement est un bien public, c'est-à-dire que les deux parties contribuent à sa production et qu'elle n'entraîne pas de rivalité dans sa consommation. Nous pouvons légitimement tracer le parallèle de cette situation avec la relation étudiant-enseignant en première année de licence.

En effet, les étudiants effectuent, dans leur majorité, leurs premiers pas à l'université. Ils ne conçoivent donc pas encore tout le fonctionnement sous-jacent à l'organisation de la licence : le contrôle de l'assiduité, la gestion de l'emploi du temps, la position particulière des enseignants (enseignant-chercheur, professeurs agrégés, doctorants, vacataires...) et le type de connaissances à assimiler. Ils sont donc en condition d'absorber le discours des enseignants pour en saisir les traits de fonctionnement qu'ils leur permettront de réussir et diminuer la peur de l'échec.

En considérant l'évaluation de la participation, les étudiants ne peuvent pas encore formuler de croyances à son égard, de même que sur la façon dont elle est évaluée, parce qu'ils ne l'ont pas encore éprouvée.

À ce niveau de la formation, les étudiants ont dans leur majorité « choisi » de venir dans ce type de cursus d'enseignement supérieur pour plusieurs raisons : l'accueil sans sélection de tous les profils de bacheliers, le temps effectif en présentiel réduit, la possibilité de réfléchir pour se réorienter, la localisation géographique, etc.

Bien que leurs attentes individuelles soient différentes, elles sont néanmoins tournées vers l'acquisition de connaissances propres à la formation et à la discipline pour avancer dans leur formation supérieure, mais également pour affiner leur choix de carrière professionnelle.

De plus, la peur de ne pas réussir renforce leur motivation à éviter cette situation.

De surcroît, le module d'introduction à l'économie est un module fondamental de leur licence dont les connaissances transmises permettront d'appréhender le reste de la formation en

économie et qui compte pour un cinquième de l'ensemble des crédits qu'il est possible d'acquérir au premier semestre. Les étudiants sont donc en attente de l'instruction que leur offrira l'équipe enseignante, laquelle conçoit par définition un intérêt à délivrer cette instruction.

Dans ce contexte, l'opacité liée à la nouveauté quant aux règles de fonctionnement, et les attentes positives sur le module de formation permettent de faire accepter aux étudiants des dispositifs qui leur semblent opportuns pour leurs stratégies de réussite à l'université. Par conséquent, le pouvoir de négociation de l'évaluation a été accepté et le processus de négociation a été validé par les étudiants.

Dans la condition expérimentale, ils sont ainsi en mesure de « jouer le jeu » puisqu'il n'y a pas d'autres conséquences que la note de participation sur laquelle ils ne savent pas encore dans quelle mesure elle aura un impact sur leur réussite.

Ils ont manifesté un effort dédié à se conformer aux règles coconstruites de la participation au module d'introduction à l'économie. Ils ont adopté un cadre d'apprentissage favorable à la réussite en développant dans ces groupes une communauté d'apprentissage, à l'instar de ce qui était observé dans les interventions de chercheurs en science de l'éducation ce qui leur a permis d'acquérir des capacités d'autonomie et de prise de décision dans leur apprentissage (Deeley, 2014). Cela s'est traduit par la visibilité d'interrelations entre les étudiants pendant les travaux dirigés en introduction à l'économie, qui se sont reproduites dans les autres travaux dirigés. Si l'on considère que ces échanges entre étudiants existent également dans les autres groupes, ils ont été plus sujets à faire progresser les étudiants pour assimiler les connaissances et les règles d'organisation de la licence.

Les dispositions qu'ils ont développées dans ce contexte auraient alors été jugées valables pour leur cursus à la fois pour le module dans lequel se déroule l'expérimentation et pour les autres modules de formation.

La conséquence de cette réception favorable a été un effort important pour atteindre de hauts scores pour le contrôle continu. La mise en pratique de ces dispositions dans un contexte non expérimental a permis d'engranger une réussite supérieure par rapport aux autres étudiants. L'exécution du contrat par l'enseignant a été crédible, c'est-à-dire que l'enseignant a tenu son engagement à utiliser la grille d'évaluation co-construite pour juger la participation en travaux dirigés et cela a engendré un effort d'apprentissage constant au cours des examens en contrôle continu. La vérifiabilité du contrat a été permise par la note de participation à la fin du processus expérimental.

L'optimum social peut être donc être atteint dans le contexte d'initialisation de la formation lorsque les étudiants et les enseignants entrent en interaction pour dialoguer et contribuer à une règle commune d'évaluation.

Cependant, lorsque la période des examens du premier semestre est arrivée, l'intensité de l'effort et l'utilisation d'aptitudes individuelles qui pouvaient naître de la condition expérimentale ont été atténuées. Ainsi, le contrôle terminal n'a-t-il pas été mieux réussi par les étudiants du groupe traité.

Il est possible d'expliquer cette absence d'effet par l'architecture du contrôle des connaissances, dans laquelle le contrôle terminal compte pour la moitié de la note finale d'une matière donnée.

De plus, il existe une part d'aléatoire dans les examens terminaux (Becker et Rosen, 1996) et cette période concentre tous les tests sur quelques jours avec des amplitudes horaires importantes.

De surcroît, c'est la première expérience de partiels pour les étudiants. Ces derniers ne bénéficient pas d'entraînement au préalable, les formes d'évaluation pouvant sensiblement différer des formes d'examens en contrôle continu.

Il serait également possible d'expliquer l'écart de notes moyennes non significatif entre le groupe traité et le groupe témoin lors du contrôle terminal par le fait que l'examen en lui-même ne permettait pas d'appliquer des dispositions acquises au cours du processus expérimental. Ces éléments peuvent provoquer un stress qui agirait négativement avec la capacité à investir davantage d'efforts en puisant dans des capacités spécifiques à l'enseignement supérieur.

Les étudiants auraient aussi considéré que leurs notes de contrôle continu étaient suffisantes pour valider le semestre et ils ont donc réduit leurs efforts pour atteindre une note moins élevée lors du contrôle terminal. En d'autres termes, il suffisait pour eux d'atteindre une note de compensation suffisante pour valider la matière. C'est ce qui a dû se produire pour le contrôle terminal d'introduction à l'économie pour lequel la note moyenne s'élève à 7/20 pour le groupe traité. Ce dernier a atteint en moyenne la note de 13/20 pour le contrôle continu. Cela expliquerait pourquoi les étudiants sous la condition expérimentale n'acquièrent pas de gain de réussite lors du partiel d'introduction à l'économie.

Néanmoins, leurs résultats ont été meilleurs que le groupe témoin sur les autres matières pendant la période d'examens en contrôle terminal. Les étudiants du groupe traité ont ainsi mieux réparti leurs efforts d'apprentissage sur toutes les matières, afin de maximiser leurs chances de valider leur semestre. Cette répartition de l'effort peut être engendrée par une meilleure prise en considération de l'acquisition de connaissances pour chaque examen et une meilleure compréhension de ce qui était attendu d'eux.

En fin de compte, cette stratégie de réussite ne s'est pas traduite par de meilleures chances de valider le premier semestre par rapport à des étudiants qui n'ont pas bénéficié du dispositif de partenariat évaluatif. Dans ce cadre, l'effort supérieur fourni lors des examens en contrôle continu n'a pas été suffisant pour compenser la baisse de l'effort aux examens terminaux. Les notes ont été meilleures, mais de façon générale la promotion a généralement validé son semestre, sans que ce dispositif n'apporte de chances supplémentaires.

b. En deuxième année, la méfiance des étudiants atténuée la portée de la négociation

En revanche, les étudiants qui sont arrivés en deuxième année ont franchi des étapes qui ont manifestement modifié leurs attentes sur la formation et leurs stratégies de réussite. Par conséquent, la relation entre les étudiants et les enseignants se transforme. Cette situation est alors plus favorable à une divergence d'intérêts quant à la formation, ce qui correspond à la deuxième manière d'appréhender d'un point de vue théorique la relation enseignant-étudiant.

Pour les étudiants de deuxième année, d'une part ils sont rompus à l'ensemble des règles régissant la formation et d'autre part ils connaissent le jeu d'acteurs de l'équipe pédagogique et administrative.

En ce qui concerne la participation, ils ont ainsi internalisé les pratiques et le discours des enseignants sur la manière de la juger en travaux dirigés et ils sont en mesure d'apprécier la validité et la portée de son évaluation. En d'autres termes, ils connaissent les problèmes d'efficacité et savent s'en accommoder dans leur cursus universitaire.

À ce stade de leurs études, les étudiants ont déjà subi une forme de sélection pour continuer leur cursus. Par conséquent, le fait d'avoir validé leur première année peut atténuer leur peur de l'échec. Ils connaissent désormais mieux leurs forces et leurs faiblesses. Leur projet de formation et leur envie d'insertion professionnelle se perfectionnent au fur et à mesure de l'avancée de la licence. Ils formulent également des préférences précises pour l'un des deux volets de la formation.

Il semble aussi nécessaire de considérer que pour la majorité des étudiants, leurs attentes et leur engagement envers le module de microéconomie sont faibles relativement à d'autres matières. Dans le discours des étudiants, c'est une discipline trop éloignée de la réalité et qui ne fournit pas des outils directement mobilisables dans un environnement professionnel. La conséquence de cette conception est un faible effort à l'apprentissage qui se traduit régulièrement par de faibles résultats aux examens, en première année et en deuxième année. Ils abandonnent délibérément ce module pour se focaliser sur les autres. Les étudiants ne le considèrent pas comme faisant partie intégrante de leur réussite dans la formation, tandis que les enseignants le considèrent comme étant une matière fondamentale de la discipline.

Ainsi, lorsqu'un enseignant leur propose de participer à un partenariat dont l'objectif est de négocier la règle d'évaluation de la participation, ils se méfient de son discours et de sa véracité, autant qu'ils doutent de son impact sur leur réussite. Alors que dans les deux groupes expérimentaux de deuxième année, les étudiants ont accepté à l'unanimité de jouer le jeu de la négociation de l'évaluation de la participation, ils n'ont pas pris en considération les apports de cette démarche dans leur processus d'apprentissage.

Dans ce cas, bien que l'expérimentation ait donné l'opportunité de négocier un contrat d'évaluation de la participation et que les étudiants aient accepté d'y participer, ces derniers n'ont pas admis que l'exécution du contrat serait crédible. La négociation n'a alors pas engendré de réduction de la croyance négative sur les règles d'évaluation et n'a pas permis de donner davantage d'intérêt à l'effort d'apprentissage dans ce module.

La condition expérimentale a ainsi conduit les étudiants à davantage rester sur leurs acquis de connaissances antérieurs pour se conformer à la règle co-construite, sans penser qu'elle se traduirait par une véritable évaluation. Pour eux, il n'y avait ainsi pas assez d'incitation à fournir un effort pour ce module, puisque, quelle que soit la quantité d'efforts fournis, cela se traduirait par de l'échec aux examens.

De plus, la proposition de l'enseignant a pu être mal interprétée par les étudiants parce qu'elle entraînait en contradiction avec leur socle de compréhension des règles de la formation. Cette nouveauté a pu engendrer un stress différent pour répondre à des attentes différentes à celles auxquelles ils étaient habitués à se conformer dans un contexte sans changement.

Par conséquent, la condition expérimentale a détérioré la performance plutôt que de la maintenir ou de la renforcer lors des deux examens en contrôle continu.

Cette nouveauté a donc été mal perçue par des étudiants dans un module pour lequel ils estimaient que, quelles que soient les conditions du changement, elles ne modifieraient pas fondamentalement leurs manières d'apprécier son contenu et ses méthodes. Au contraire, cela les forçait à davantage s'impliquer dans un module où l'effort d'implication n'est de toute manière pas récompensé par un résultat positif.

De plus, le fait que l'exécution du contrat n'ait pas été rendue visible avant la fin du semestre a pu conforter les étudiants dans leurs croyances négatives.

En revanche, lorsque nous analysons le résultat au partiel, l'effet est inversé : les étudiants ont mieux réussi au contrôle terminal de microéconomie et ont également globalement mieux réussi aux examens terminaux de tous les autres modules.

Ce renversement de situation peut s'expliquer par la prise en compte des étudiants de leur note de participation au module, c'est-à-dire la fin du processus expérimental. En effet, l'issue du partenariat se conclut par la note qui est calculée à partir des critères co-déterminés par le groupe d'étudiants et par l'enseignant. En moyenne, la note de participation obtenue par le groupe expérimental est de 13,88 ($\sigma^2 = 1,84$) ce qui correspond environ au double de la note obtenue lors des examens en contrôle continu. Le rendu de la note grâce à la distribution de la grille et des

commentaires associés à l'évaluation de la participation a été vécu positivement par les étudiants qui ne s'attendaient pas à ce type de résultats. Pour les étudiants qui ont mesuré l'ampleur de l'écart entre leurs résultats aux tests de connaissances et à cette évaluation de leurs autres aptitudes, cela a produit un changement de comportement important. En tenant son engagement de respect du partenariat, l'enseignant a ainsi conduit les étudiants à se remettre en question sur leurs stratégies de réussite.

Cette note aurait ainsi produit un changement de comportement dans l'effort d'acquisition de connaissances, aussi bien en microéconomie, pour tenter de rattraper les résultats du contrôle continu, mais également pour les autres examens. Cette expérience a permis de leur montrer qu'ils avaient les capacités de réussir et qu'il ne fallait pas demeurer dépendant au sentier de leur apprentissage, mais qu'ils devaient prendre en main leur parcours d'études. Leur effort d'apprentissage s'est soldé par une réussite plus importante que le groupe témoin aux examens terminaux.

In fine, cette remise en question a permis d'augmenter les chances de valider le troisième semestre pour les étudiants du groupe expérimental. Par conséquent, la note correspondant au partenariat évaluatif, la prise en compte des aptitudes acquises dans la formation pour réussir et les dispositions qui étaient offertes par le dispositif expérimental ont permis à ces étudiants de reprendre confiance en eux au cours de leur troisième semestre d'études et de remettre en cause le processus d'échec qui semblait inévitable. Ce processus s'est traduit par davantage de réussite.

À la différence des étudiants de première année, c'est donc davantage la manière dont l'issue de la négociation s'est réalisée qui aurait fait changer les comportements sur l'effort d'apprentissage et sur la capacité de réussir aux examens que le fait de se saisir de cette négociation pour adopter des dispositions favorables à la réussite. Lorsque le contrat a été exécuté par l'enseignant, c'est-à-dire qu'il a tenu l'engagement pris avec les étudiants de tenir compte de leurs critères d'évaluation pour évaluer leur participation, les étudiants ont alors pris conscience de la véracité du contrat et se sont remis en question. Finalement, cette remise en question a permis aux étudiants de renforcer leur motivation et de croire en leurs capacités y compris dans un module sur lequel ils ne portaient pas d'intérêt pour leur réussite et ils ont donc fourni un effort plus important lors du contrôle terminal de connaissances.

Ainsi, l'expérimentation permet de compléter les apports de la littérature sur la négociation dans un cadre d'aléa moral en fournissant l'idée selon laquelle l'issue de la négociation (le contrat) doit être visible par les parties prenantes afin de rendre crédible l'accord et de réduire les croyances négatives sur l'autre partie. Nous rentrons ainsi dans une dimension dynamique de la négociation des contrats et de la vérification de l'effort en présence de divergences d'intérêts.

Enfin, nous avons explicité que le contexte des interventions en science de l'éducation était généralement favorable à leurs implémentations. Nous montrons ainsi que lorsqu'il n'est pas favorable, cela peut conduire à des effets non anticipés et qu'il peut détériorer la réussite. Par conséquent, la négociation des règles d'évaluation produirait un meilleur effet en début de processus de formation lorsque les attentes des deux parties de la relation sont sur le même niveau, c'est-à-dire la contribution à un bien commun, parce qu'elle permet de comprendre les règles en vigueur et de promouvoir des environnements plus propices à la parole des étudiants pour améliorer ce que leurs enseignants proposent d'évaluer. Lorsque la négociation est implantée au cours du processus de formation, une condition d'efficacité serait de rendre visible l'exécution du contrat pour réduire les croyances négatives.

9. CONCLUSION

Nous avons conduit un partenariat expérimental d'évaluation dans le cadre de deux modules de formation de la licence d'Économie-Gestion de l'Université Paris-Est Marne-la-Vallée. Ce partenariat d'évaluation consistait à négocier les critères d'appréciation de la participation des étudiants dans les travaux dirigés. Nous avons mesuré l'impact de ce processus sur les notes obtenues dans le module et sur les résultats dans les autres composantes de la formation. Nous montrons que la négociation produit un effet positif sur la réussite quand la relation entre l'enseignant et les étudiants est favorable à la négociation. L'effet positif sur la réussite dépend également de la visibilité de l'exécution du contrat dans le cas où la relation entre les enseignants et les étudiants serait constituée de divergences d'intérêt.

Cette expérimentation répond en partie à des préoccupations générales sur le modèle évaluatif de l'université. En reprenant les termes de l'Inspection générale de l'administration de l'éducation nationale et de la recherche (Gauthier et al., 2007), il s'agit de promouvoir une matrice évaluative, plus plastique en fonction des publics accueillis et des acquis visés, qui permet de construire l'évaluation concomitamment avec l'enseignement et qui s'oriente vers une évaluation des compétences en complément des connaissances.

Nous contribuons ainsi à une démarche générale d'expérimentations de dispositifs organisationnels innovants qui permet de tester les conditions de leur mise en œuvre à une échelle globale et leurs impacts à divers degrés (Shaw, 2009).

D'une part, nous apportons une contribution à l'ensemble des études de mesures d'impact de dispositifs de remédiation dans l'enseignement supérieur (Sneyers et de Witte, 2017). Nous complétons ainsi le panorama des outils à la disposition des responsables de formation et des enseignants pour améliorer la conduite des étudiants dans leur parcours de formation, à l'instar du tutorat, des bourses de mobilité ou sur critères sociaux, des enseignements de méthodologie, des cours de renforcement, des règles de compensation d'unité d'enseignement, etc. Nous développons l'idée selon laquelle l'accompagnement des étudiants, présentant des profils et un passé éducatif de plus en plus hétérogènes, doit s'outiller de plusieurs techniques que les acteurs de la formation peuvent utiliser.

Dans ce cadre, nous renforçons l'idée que l'un des déterminants de l'engagement des étudiants dans leur apprentissage se noue dans la relation qu'ils créent avec l'équipe enseignante (Correa et Gruver, 1987). Ces interactions doivent s'incarner au travers de dialogues et de coopérations entre les deux pôles de la relation, afin de promouvoir la parole et les décisions des étudiants qui sont toutes aussi valables que celles des enseignants (Jarvis, Dickerson et Stockwell, 2013). Par conséquent, le poids attribué à l'enseignant dans la formation peut être restreint afin de favoriser l'émergence de dispositions nouvelles de la part des étudiants, dispositions que nous avons citées à plusieurs reprises comme étant constitutives de l'apport de la formation supérieure : l'autonomie, la prise de décision, la réflexion critique.

D'autre part, nous montrons qu'il est possible d'expérimenter des outils innovants pour améliorer les pratiques d'évaluation qui soutiennent l'effort d'apprentissage davantage qu'elles ne sanctionnent sa progression (McMillan, 2015). L'inertie de l'évaluation à l'œuvre à l'université n'est ainsi pas immuable, elle peut au contraire être renversée pour incorporer plus de considérations « formatives ». Il existe des marges de progression et les organisations d'enseignement supérieur peuvent s'appuyer sur le partenariat évaluatif pour renforcer la fiabilité et la validité des examens. Dans ce contexte, l'évaluation peut être transformée et devenir dépendante des préconisations formulées aussi bien par les étudiants que par les enseignants. Une

meilleure technique d'évaluation et de contrôle des connaissances et des compétences permet d'obtenir une meilleure qualité du signal émis sur la production des diplômes.

Le message porté par le partenariat évaluatif est que la conception des tests doit aussi faire partie intégrante de la conception de l'enseignement.

D'un point de vue de microéconomie des organisations, nous apportons des arguments favorables à la mise en place d'environnements de négociation des dispositifs de vérification de l'information. Cette négociation procure des marges de manœuvre qui incitent à l'effort pour atténuer l'aléa moral sur les positions individuelles. Il est possible de donner du pouvoir à l'agent sans qu'il opte pour un comportement opportuniste et de réduire le pouvoir du principal afin qu'il n'abuse pas de sa position dominante. Cet environnement de négociation peut aboutir à des contrats qui sont constitutifs d'une allocation optimale des ressources entre les parties de la relation.

La recherche que nous avons conduite n'est cependant pas exempte d'un certain nombre de limites qui peut réduire la pertinence de son message.

Premièrement, les caractéristiques de l'expérimentation soulèvent un problème de validité externe. La taille des effectifs, qui est faible en comparaison d'études de plus grande envergure (par exemple Dmitrijeva et al., 2014), engendre nécessairement des précautions quant à la volonté de généraliser à l'ensemble de la population étudiante les effets que nous mesurons. Le fait qu'elle ait été réalisée au sein d'une unique unité de formation et de recherche, dans une seule université, restreint également la portée des résultats que nous pourrions obtenir dans un environnement de formation différent qui comporte des modalités de fonctionnement distinctes, voire opposées. Nous rappelons également que les partenariats évaluatifs ont été menés avec un seul enseignant et que par conséquent la réplication de ces processus demeure conditionnée par les pratiques et les croyances individuelles sur les normes d'évaluation à adopter, c'est-à-dire que dans ce contexte il serait également important de mettre en lumière les dispositions des enseignants à réaliser la négociation. Enfin, la temporalité du dispositif pose une question sur la durabilité des effets constatés.

Deuxièmement, la condition expérimentale est un processus défini comme la suite continue d'opérations aboutissant à un résultat davantage que l'intervention d'un procédé unique. Par conséquent, il existe des impacts sous-jacents engendrés par la condition expérimentale que nous ne sommes pas en mesure d'appréhender. Ainsi, en l'absence d'une modélisation plus fine de ce processus, nous avons délibérément évacué des dimensions qui renforcent ou atténuent l'impact sur les indicateurs de réussite. Par exemple, nous n'avons pas tenu compte du degré d'autonomie développé par les étudiants grâce à ce dispositif. Nous avons considéré que cette disposition s'était davantage produite chez le groupe expérimental et qu'elle permettait de générer un plus grand effort d'apprentissage et une réussite plus importante. La complexité du traitement et ses interactions avec d'autres variables sont par conséquent masquées.

Cette remarque nous permet également de parvenir à l'idée selon laquelle le phénomène de l'échec / de la réussite est protéiforme et que toutes ses dimensions n'ont pas été investiguées. Il reste un enjeu important de formalisation de la réussite qu'il ne faudrait pas réduire à deux indicateurs. Il serait donc nécessaire d'identifier d'autres variables de résultats pour apprécier les effets de la négociation des règles de l'évaluation dans leur ensemble.

Dans ce cadre, cette étude souffre de ne pas proposer de comparaison de l'issue des partenariats, à savoir le « contrat » contenant les règles d'évaluation et les notes obtenues par les étudiants, avec ce qui se pratique dans les autres modules de formation. Nous n'en avons pas tenu compte,

car il n'était pas possible de comparer les niveaux de participation obtenus par les étudiants des autres groupes de travail dirigés en raison d'échelles de mesure différentes et de critères de jugements variés. Par conséquent, la confrontation des résultats de participation dans les deux groupes d'étudiants n'aurait pas permis de produire des interprétations quant au niveau effectif de participation.

Enfin, une limite de cette recherche tient à la nature des partenariats offerts aux étudiants. Deeley (2014) a en effet déterminé que les coopérations peuvent être fortes, lorsque les étudiants et les enseignants conviennent d'une règle d'évaluation et/ou réalisent conjointement l'évaluation, ou faibles, lorsque les critères d'examen ne sont pas coconstruits et/ou que l'évaluation n'est pas conduite par les deux parties. Dans notre protocole expérimental, le partenariat correspond davantage au modèle « faible », dans le sens où les étudiants n'ont pas utilisé la grille pour s'autoévaluer sur la participation, l'évaluation restant sous la responsabilité du chargé de travaux dirigés. L'engagement des étudiants n'allait donc pas au-delà de la participation à la discussion et à la formalisation des critères de notation. De plus, la négociation n'a pas concerné les examens de contrôle des connaissances, pour lesquels les enjeux sont plus importants : ils structurent davantage le cursus de la formation parce qu'ils font partie intégrante du rituel de l'apprentissage et récompensent davantage l'effort des étudiants. Par conséquent, la pertinence de la négociation portant sur la participation pourrait être remise en question.

L'ensemble de ces remarques permet de dresser de nouveaux jalons qui offriraient des perspectives inédites en vue de futures recherches. Dans la mesure où nous pouvons considérer cette expérimentation comme une invitation à d'autres initiatives expérimentales, il serait judicieux de changer d'échelle afin d'incorporer un plus grand nombre de parties prenantes et produire de nouveaux résultats.

Il est souhaitable d'approfondir l'idée de la négociation de l'évaluation dans un cadre plus ambitieux, par exemple au niveau d'un examen de connaissances en contrôle continu, voire en contrôle terminal. Tout l'enjeu réside alors dans la délimitation de la marge de négociation des parties prenantes pour ne pas amoindrir le caractère certificatif de l'évaluation des connaissances des étudiants et ne pas engendrer des comportements que les responsables de formation voudraient éviter. L'objectif serait de continuer à développer les aptitudes et les compétences issues de la formation supérieure en symbiose avec l'objectif d'acquisition de connaissances disciplinaires.

Une voie ouverte de recherche serait de quantifier d'autres variables de résultats à l'issue de ces dispositifs. Ces variables seraient en phase avec l'idée du développement d'aptitudes propres à l'enseignement supérieur et pourraient ainsi s'inscrire durablement dans une économie de la connaissance qui mobilise les facultés d'apprentissage et d'évaluation individuelle tout au long de la vie.

Dans le but d'approfondir la notion de négociation de la règle d'évaluation, il serait judicieux de déterminer dans quelle mesure les différentes formes de contrats qui découlent de ces négociations influencent à la hausse ou à la baisse les indicateurs d'intérêt. Ainsi, il serait possible de proposer des choix d'évaluation, sous la forme de menu, dans lesquels les étudiants se positionneraient. À l'instar du choix du module, les étudiants choisiraient leurs évaluations.

Enfin, au moment où la loi relative à l'orientation et à la réussite des étudiants dans les universités instaure l'obligation de signature d'un contrat pédagogique liant l'étudiant à son cursus de formation, dans quelle mesure l'incorporation au sein de ces contrats de stipulations liées à l'évaluation renforcerait-elle leur crédibilité et permettrait-elle de faire adopter des dispositions individuelles nouvelles ?

BIBLIOGRAPHIE

- Abdelmalak, M. M. M. (2016). Faculty-Student Partnerships in Assessment. *International Journal of Teaching and Learning in Higher Education*, 28(2), 193-203.
- Aina, C., Baici, E., Casalone, G., & Pastore, F. (2018). *The Economics of University Dropouts and Delayed Graduation: A Survey* (Discussion Paper No. 11421). IZA Institute of Labor Economics.
- Arnold, I. J. M., & Straten, J. T. (2012). Motivation and Math Skills as Determinants of First-Year Performance in Economics. *The Journal of Economic Education*, 43(1), 33-47. <https://doi.org/10.1080/00220485.2012.636709>
- Balkenborg, D. (2001). How Liable Should a Lender Be? The Case of Judgment-Proof Firms and Environmental Risk: Comment. *American Economic Review*, 91(3), 731-738. <https://doi.org/10.1257/aer.91.3.731>
- Becker, W. E., & Rosen, S. (1992). The learning effect of assessment and evaluation in high school. *Economics of Education Review*, 11(2), 107-118. [https://doi.org/10.1016/0272-7757\(92\)90002-K](https://doi.org/10.1016/0272-7757(92)90002-K)
- Bellity, E., Gilles, F., & L'Horty, Y. (2017). *Does practicing literacy skills improve academic performance in first-year university students? Results from a randomized experiment* (Working Paper No. 2017-02). Travail, emploi et politiques publiques.
- Birenbaum, M. (1996). Assessment 2000: Towards a Pluralistic Approach to Assessment. In M. Birenbaum & F. J. R. C. Dochy (Éd.), *Alternatives in Assessment of Achievements, Learning Processes and Prior Knowledge* (p. 3-29). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-011-0657-3_1
- Bishop, J. (2006). Chapter 15: Drinking from the Fountain of Knowledge: Student Incentive to Study and Learn – Externalities, Information Problems and Peer Pressure. In *Handbook of the Economics of Education* (Vol. 2, p. 909-944). Elsevier. [https://doi.org/10.1016/S1574-0692\(06\)02015-0](https://doi.org/10.1016/S1574-0692(06)02015-0)
- Black, P., & Wiliam, D. (1998). Assessment and Classroom Learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7-74. <https://doi.org/10.1080/0969595980050102>
- Bloom, B. S., Hastings, T. J., & Madaus, G. F. (1971). *Handbook on formative and summative evaluation of student learning*. New York: McGraw-Hill.
- Bloxham, S., den-Outer, B., Hudson, J., & Price, M. (2016). Let's stop the pretence of consistent marking: exploring the multiple limitations of assessment criteria. *Assessment & Evaluation in Higher Education*, 41(3), 466-481. <https://doi.org/10.1080/02602938.2015.1024607>
- Bonesrønning, H. (2004). Do the teachers' grading practices affect student achievement? *Education Economics*, 12(2), 151-167. <https://doi.org/10.1080/0964529042000239168>
- Bonesrønning, H., & Opstad, L. (2015). Can student effort be manipulated? Does it matter? *Applied Economics*, 47(15), 1511-1524. <https://doi.org/10.1080/00036846.2014.997923>
- Bovill, C., Cook-Sather, A., & Felten, P. (2011). Students as co-creators of teaching approaches, course design, and curricula: implications for academic developers. *International Journal for Academic Development*, 16(2), 133-145. <https://doi.org/10.1080/1360144X.2011.568690>
- Brown, B. W., & Saks, D. H. (1981). The Microeconomics of Schooling. *Review of Research in Education*, 9(1), 217-254. <https://doi.org/10.3102/0091732X009001217>
- Brown, B. W., & Saks, D. H. (1987). The microeconomics of the allocation of teachers' time and student learning. *Economics of Education Review*, 6(4), 319-332. [https://doi.org/10.1016/0272-7757\(87\)90015-X](https://doi.org/10.1016/0272-7757(87)90015-X)

- Brown, S., & Knight, P. (1994). *Assessing learners in higher education*. London ; Philadelphia: Kogan Page.
- Correa, H., & Gruver, G. W. (1987). Teacher-student interaction: A game theoretic extension of the economic theory of education. *Mathematical Social Sciences*, 13(1), 19-47. [https://doi.org/10.1016/0165-4896\(87\)90057-6](https://doi.org/10.1016/0165-4896(87)90057-6)
- Deeley, S. J. (2014). Summative co-assessment: A deep learning approach to enhancing employability skills and attributes. *Active Learning in Higher Education*, 15(1), 39-51. <https://doi.org/10.1177/1469787413514649>
- Deeley, S. J., & Bovill, C. (2017). Staff student partnership in assessment: enhancing assessment literacy through democratic practices. *Assessment & Evaluation in Higher Education*, 42(3), 463-477. <https://doi.org/10.1080/02602938.2015.1126551>
- Deeley, S. J., & Brown, R. A. (2014). *Learning Through Partnership in Assessment* (Teaching and learning together in higher education No. 13). Bryn Mawr College.
- Demougin, D., & Helm, C. (2006). Moral Hazard and Bargaining Power. *German Economic Review*, 7(4), 463-470. <https://doi.org/10.1111/j.1468-0475.2006.00130.x>
- Dittrich, M., & Städter, S. (2015). Moral hazard and bargaining over incentive contracts. *Research in Economics*, 69(1), 75-85. <https://doi.org/10.1016/j.rie.2014.10.002>
- Dmitrijeva, J., Du Parquet, L., L'Horty, Y. & Petit, P. (2015). 7. L'échec en licence pour cause d'excès de travail... salarié. *Regards croisés sur l'économie*, 16(1), 117-130. doi:10.3917/rce.016.0117.
- Doyle, E., Buckley, P., & Whelan, J. (2018). Assessment co-creation: an exploratory analysis of opportunities and challenges based on student and instructor perspectives. *Teaching in Higher Education*, 1-16. <https://doi.org/10.1080/13562517.2018.1498077>
- Duguet, A., Le Mener, M., & Morlaix, S. (2016). The Key Predictors of Success in University in France: What Are the Contributing Factors and Possible New Directions in Educational Research? *International Journal of Higher Education*, 5(3), 222-235. <https://doi.org/10.5430/ijhe.v5n3p222>
- Falchikov, N. (1986). Product comparisons and process benefits of collaborative peer group and self-assessments. *Assessment & Evaluation in Higher Education*, 11(2), 146-166. <https://doi.org/10.1080/0260293860110206>
- Frey, B. S. (1993). Does monitoring increase work effort? The rivalry with trust and loyalty. *Economic Inquiry*, 31(4), 663-670. <https://doi.org/10.1111/j.1465-7295.1993.tb00897.x>
- Garside, J., Nhemachena, J. Z. Z., Williams, J., & Topping, A. (2009). Repositioning assessment: Giving students the 'choice' of assessment methods. *Nurse Education in Practice*, 9(2), 141-148. <https://doi.org/10.1016/j.nepr.2008.09.003>
- Gauthier, R.-F., Caffin-Ravier, M., Descamps, B., Mosnier, M., & Peretti, H. (2007). *L'évaluation des étudiants à l'Université: point aveugle ou point d'appui?* (No. 2007-072). Inspection générale de l'administration de l'Éducation nationale et de la Recherche.
- Gibbs, G., & Simpson, C. (2004). Conditions Under Which Assessment Supports Students' Learning. *Learning and Teaching in Higher Education*, 1, 3-31.
- Gillet, I., Greenan, N., & Le Gall, R. (2016). *Les effets de la surveillance électronique : une expérimentation dans un centre d'appels*. Paris: Lavoisier.
- Hölmstrom, B. (1979). Moral Hazard and Observability. *The Bell Journal of Economics*, 10(1), 74-91. <https://doi.org/10.2307/3003320>

- Jarvis, J., Dickerson, C., & Stockwell, L. (2013). Staff-student Partnership in Practice in Higher Education: The Impact on Learning and Teaching. *Procedia - Social and Behavioral Sciences*, 90, 220-225. <https://doi.org/10.1016/j.sbspro.2013.07.085>
- Kalai, E., & Smorodinsky, M. (1975). Other Solutions to Nash's Bargaining Problem. *Econometrica*, 43(3), 513-518. <https://doi.org/10.2307/1914280>
- L'Horty, Y., & Petit, P. (2011). Évaluation aléatoire et expérimentations sociales. *Revue française d'économie*, XXVI(1), 13-48. <https://doi.org/10.3917/rfe.111.0013>
- Leslie, L. J., & Gorman, P. C. (2017). Collaborative design of assessment criteria to improve undergraduate student engagement and performance. *European Journal of Engineering Education*, 42(3), 286-301. <https://doi.org/10.1080/03043797.2016.1158791>
- McKeachie, W. J. (1962). Procedures and Techniques of Teaching: A Survey of Experimental Studies. In N. Sanford (Éd.), *The American college: A psychological and social interpretation of the higher learning*. (p. 312-364). Hoboken: John Wiley & Sons Inc. <https://doi.org/10.1037/11181-008>
- McMillan, J. H. (Éd.). (2013). *Handbook of research on classroom assessment*. Los Angeles: Sage Publications.
- Michaelis, J., & Schwanebeck, B. (2016). Examination rules and student effort. *Economics Letters*, 145, 65-68. <https://doi.org/10.1016/j.econlet.2016.05.019>
- Morlaix, S., & Lambert-Le Mener, M. (2015). La motivation des étudiants à l'entrée à l'université : quels effets directs ou indirects sur la réussite? *Recherches en Éducation*, (22), 152-167.
- Morlaix, S., & Perret, C. (2013). L'évaluation du Plan Réussite en Licence : quelles actions pour quels effets ? *Recherches en Éducation*, (15), 137-150.
- Nash, J. F. (1950). The Bargaining Problem. *Econometrica*, 18(2), 155-162. <https://doi.org/10.2307/1907266>
- OECD (Éd.). (2013). *Synergies for better learning: an international perspective on evaluation and assessment*. Paris: OECD.
- Pitchford, R. (1998). Moral hazard and limited liability: The real effects of contract bargaining. *Economics Letters*, 61(2), 251-259. [https://doi.org/10.1016/S0165-1765\(98\)00141-4](https://doi.org/10.1016/S0165-1765(98)00141-4)
- Pratt, J. W., & Zeckhauser, R. (Éd.). (1985). *Principals and agents: the structure of business*. Boston, Mass: Harvard Business School Press.
- Razafindratsima N., (2018), Parcours et réussite en licence et en Paces : les résultats de la session 2017, Note Flash n°21, MESRI - Service d'information et études statistiques (SIES).
- Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, Schools, and Academic Achievement. *Econometrica*, 73(2), 417-458. <https://doi.org/10.1111/j.1468-0262.2005.00584.x>
- Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5), 688-701. <https://doi.org/10.1037/h0037350>
- Rust, C., Price, M., & O'Donovan, B. (2003). Improving Students' Learning by Developing their Understanding of Assessment Criteria and Processes. *Assessment & Evaluation in Higher Education*, 28(2), 147-164. <https://doi.org/10.1080/02602930301671>
- Scriven, M. (1967). The methodology of evaluation. In R. Tyler W., R. Gagné, & M. Scriven (Éd.), *Perspectives of curriculum evaluation* (Rand McNally, p. 39-83). Chicago.
- Shaw, K. (2009). Insider econometrics: A roadmap with stops along the way. *Labour Economics*, 16(6), 607-617. <https://doi.org/10.1016/j.labeco.2009.09.001>

- Sneyers, E., & De Witte, K. (2018). Interventions in higher education and their effect on student success: a meta-analysis. *Educational Review*, 70(2), 208-228. <https://doi.org/10.1080/00131911.2017.1300874>
- Stefani, L. A. J. (1998). Assessment in Partnership with Learners. *Assessment & Evaluation in Higher Education*, 23(4), 339-350. <https://doi.org/10.1080/0260293980230402>
- Zubrickas, R. (2015). Optimal grading. *International Economic Review*, 56(3), 751-776. <https://doi.org/10.1111/iere.12121>

Annexes

a. Réussite à l'université en France

Nous avons compilé plusieurs rapports d'information du Ministère de l'Enseignement supérieur pour dresser l'évolution du taux de diplomation en trois années en licence, depuis la généralisation du diplôme en 2004. Les dernières statistiques descriptives disponibles correspondent à la promotion d'étudiants de l'année 2012-2013 pour laquelle il est possible de déterminer son devenir trois années ou quatre années après la première inscription. Les indicateurs les plus révélateurs du phénomène de la réussite sont inscrits dans le tableau A1.

Ces indicateurs n'offrent qu'une vision partielle de la réussite à l'université parce que d'une part, ils ne tiennent compte que des seules réorientations au sein de l'université et d'autre part, ils sont calculés sur la base des inscriptions administratives. Par conséquent, les réorientations vers des filières autres qu'universitaires (paramédicale ou sociale, école d'ingénieurs...) et la présence effective des étudiants en cours de formation ne sont pas utilisées.

De plus, chaque université possède sa propre population étudiante, c'est-à-dire que les caractéristiques sociodémographiques et les parcours scolaires antérieurs des étudiants diffèrent selon les établissements. Or, la probabilité de réussite (ou de passage) apparaît corrélée avec des caractéristiques telles que l'origine sociale et le parcours scolaire antérieur (série du baccalauréat, âge au moment de son obtention, ancienneté de son obtention, la mention obtenue). Les caractéristiques non observables de chaque établissement, par exemple les modalités d'évaluation et les dispositifs d'accompagnement des étudiants sont également des facteurs qui influencent la réussite.

Tableau A1 : Indicateurs de réussite d'obtention de la licence en trois ans

Promotion	2004-2005	2005-2006	2006-2007	2007-2008	2008-2009	2009-2010	2010-2011	2011-2012	2012-2013	2013-2014
Taux de licenciés en 3 ans	28,8*	27,9	27,8	27	27,2	26,5	27,8	27,2	28,4	28
Part de femmes	62	n.c.	n.c.	62,2	61,6	60,3	59,7	59,3	58,9	58,7
Part de bacheliers Général	73,1	n.c.	n.c.	75,3	74,2	71,4	71	69,2	75,3	77,3
Part de bacheliers Technologique	16,9	n.c.	n.c.	16,1	n.c.	n.c.	n.c.	n.c.	14,7	14,7
Part de bacheliers Professionnel	4	n.c.	n.c.	4,2	n.c.	n.c.	n.c.	n.c.	10	8
Taux de femmes licenciées en 3 ans	33	n.c.	n.c.	30,6	30,6	30,1	31,7	31,1	32,9	33,4
Taux d'hommes licenciés en 3 ans	21,9	n.c.	n.c.	21,1	21,7	21,2	21,9	21,6	21,8	21,6
Taux de bacheliers Général licenciés en 3 ans	36	n.c.	n.c.	33	33,3	33,2	34,7	35	35,8	35,1
Taux de bacheliers Technologique licenciés en 3 ans	7,4	n.c.	n.c.	9,8	7,5	8,7	9,8	9,2	8,1	7,9
Taux de bacheliers Professionnel licenciés en 3 ans	2,5	n.c.	n.c.	2,7	3,1	3	3,7	3,1	2,3	2,3
Redoublement de la première année	24,6	n.c.	n.c.	24,1	25	n.c.	n.c.	n.c.	n.c.	n.c.
Non-réinscription après la première année	27,8	n.c.	n.c.	31,8	30	n.c.	n.c.	n.c.	n.c.	n.c.
Nombre de néo-entrants à l'université	179300	172200	166100	154300	150100	n.c.	n.c.	n.c.	n.c.	153000†
Nombre de néo-bacheliers										

Note : * parmi les 179300 néo-entrants à l'université lors de l'année 2004-2005, 28,8 d'entre eux ont obtenu une licence trois années plus tard. n.c : non communiqué.

† : néo-bacheliers.

Sources :

- Fouquet S., (2013), Parcours et réussite en licence et en master à l'université, Note d'information n°13-02, MESR-DGESIP-DGRI, Service d'information et études statistiques (SIES).
- Fouquet S., (2014), Parcours et réussite aux diplômes universitaires : les indicateurs de la session 2012, Note Flash n°08-04, DGRI-SCSER, Service d'information et études statistiques (SIES).
- Maetz I., (2015), Parcours et réussite aux diplômes universitaires : les indicateurs de la session 2013, Note Flash n°07-05, MENESR DGESIP/DGRI - SCSESR - Service d'information et études statistiques (SIES).
- Maetz I., (2016), Parcours et réussite aux diplômes universitaires : les indicateurs de la session 2014, Note Flash n°01, MENESR DGESIP/DGRI - SCSESR - Service d'information et études statistiques (SIES).
- Maetz I., (2016), Parcours et réussite aux diplômes universitaires : les indicateurs de la session 2015, Note Flash n°15, MENESR DGESIP/DGRI - SCSESR - Service d'information et études statistiques (SIES).
- Maetz I., (2017), Parcours et réussite en licence et en Paces : les résultats de la session 2016, Note Flash n°18, MESRI - Service d'information et études statistiques (SIES).
- Razafindratsima N., (2018), Parcours et réussite en licence et en Paces : les résultats de la session 2017, Note Flash n°21, MESRI - Service d'information et études statistiques (SIES).
- Prouteau D., (2009), Parcours et réussite en licence des inscrits en L1 en 2004, Note d'information n°09-23, MESR-DGESIP-DGRI, Service d'information et études statistiques (SIES).

b. Grille d'évaluation de la participation à l'issue de la négociation

Nous présentons à titre d'illustration du protocole expérimental, un exemple de grille d'évaluation qui a été formulée à l'issue des trois séances de réflexion collective sur les critères de notation de la participation. Ces grilles ont été communiquées aux étudiants à la quatrième séance à l'issue de la validation pour qu'ils en conservent un exemplaire tout au long du semestre. L'évaluation finale de la participation a été transmise aux étudiants en entourant chaque item d'évaluation et en communiquant la note finale qui tenait compte des critères. Pour rappel, chaque item endossait la même pondération.

Grille d'évaluation de la note de participation en Travaux dirigés Introduction à l'Économie						
GROUPE de TD n° XXX						
Nom : XXX						
Règles		Fréquence			Commentaire	
Étudiants	Poser des questions (y compris si l'on pense que ce n'est pas cohérent)	JAMAIS	PARFOIS	RÉGULIÈREMENT		
	<ul style="list-style-type: none"> - Sur les mécanismes - Interprétations 					
	Vocabulaire					
	Réponse aux questions même si on pense avoir faux	JAMAIS	PARFOIS	RÉGULIÈREMENT		
	Être présent	JAMAIS	PARFOIS	RÉGULIÈREMENT		
	Être ponctuel	JAMAIS	PARFOIS	RÉGULIÈREMENT		
Chargé de TD	Être attentif.ve	JAMAIS	PARFOIS	RÉGULIÈREMENT		
	Être polis	JAMAIS	PARFOIS	RÉGULIÈREMENT		
	Apporter son matériel de travail (résumé, texte, ordinateur, notes de cours ...)	JAMAIS	PARFOIS	RÉGULIÈREMENT		
	Travailler pendant le TD (prise de notes – prise de la correction)	JAMAIS	PARFOIS	RÉGULIÈREMENT		
	Prises d'initiatives pendant le TD (bien se sentir en TD)	JAMAIS	PARFOIS	RÉGULIÈREMENT		
	Participer au débat / Réagir aux interventions de ses camarades	JAMAIS	PARFOIS	RÉGULIÈREMENT		
Interaction avec le chargé de TD pour approfondir						
Note :						

c. Test d'équivalence des notes de participation

Afin de nous assurer que la condition expérimentale a été homogène entre les groupes expérimentaux, nous recourons à une ANOVA. Ce test permet de détecter si une note moyenne est significativement différente des autres. Nous utilisons la statistique de Fisher. Le test ne permet pas de rejeter l'hypothèse de nullité des écarts entre toutes les notes moyennes. Par conséquent, il n'y a pas eu de surnotation ou de sous-notation de la participation à l'issue de la négociation de la règle d'évaluation.

Tableau A2 : Test d'homogénéité des notes moyennes de participation des groupes traités

Groupe	Nombre d'observations	Note de participation moyenne	Test de Fisher
Groupe traité n°1 L1	29	12,972 (2,028)	1,86 p-value = 0,1204
Groupe traité n°2 L1	31	12,710 (2,132)	
Groupe traité n°3 L1	26	13,077 (3,068)	
Groupe traité n°4 L2	32	13,945 (2,225)	
Groupe traité n°5 L2	29	13,822 (1,382)	

Note : les erreurs-types sont entre parenthèses. Le test de Fisher teste l'hypothèse nulle de l'existence d'une différence entre toutes les moyennes des notes de participation. Les moyennes sont calculées sur tous les étudiants ayant été dans la condition expérimentale.