

Communication — Céreq Échanges, actes des JDL 2020
Axe 3. « Les processus de sélection : le point de vue des acteurs »

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

Rémi LE GALL¹

En 2018, l'indicateur de réussite à l'université, diffusé par le Ministère de l'Enseignement Supérieur, de la Recherche et de l'Innovation, montrait que moins de trois étudiants sur dix décrochaient leur licence en trois ans parmi les néo-bacheliers inscrits en 2014-2015 (Razafindratsima & Bonnevialle, 2019). Depuis 2002, cet indicateur est stable et il ne connaît pas d'inflexion à la hausse.

Alors que le faible taux de réussite était considéré aussi bien comme un gage d'excellence du diplôme délivré qu'un signe de l'échec personnel des étudiants (Aina et al., 2018), la massification de l'enseignement supérieur a redéfini son interprétation. De nombreux dispositifs de remédiation ont été mis en œuvre pour contrer cette tendance, aussi bien au niveau national, qu'au niveau local. Les recherches entreprises pour étudier les effets de ces outils montrent le plus souvent que ces instruments sont facultatifs et touchent moins ceux qui en auraient le plus besoin (Morlaix et Perret, 2013 ; Morlaix et Lambert-Le Mener, 2015). Ainsi, les universités paraissent accuser un retard dans l'adaptation des parcours pour accompagner le flux de « nouveaux publics ». Désormais, le phénomène de l'échec remet en question la qualité de l'organisation du cursus de formation.

Pour rendre compte des tensions organisationnelles qui perdurent à l'université, nous mobilisons la relation enseignant-étudiant (Correa et Gruver, 1987), qui, sous plusieurs hypothèses, peut se caractériser comme une relation d'agence (Zubrickas, 2016), dans laquelle s'inscrivent les dispositifs d'évaluation des connaissances. Ils jouent le rôle de dispositifs de vérification de l'information (Le Gall, 2018), c'est-à-dire qu'ils collectent, enregistrent et analysent l'effort d'apprentissage fourni par les étudiants. La création d'un examen est à la discrétion de l'enseignant, qui conçoit les critères, les règles et les pondérations de l'évaluation. Ce rôle lui confère une autorité et un pouvoir de contrôle sur l'apprentissage des étudiants, à l'instar des modes de surveillance d'une relation d'agence générique (Hölmstrom, 1979). Cependant, cette dimension de la profession entre en conflit avec d'autres fonctions professionnelles et le résultat de l'arbitrage vis-à-vis du temps dévolu à l'évaluation conduit à choisir des modalités d'examens les plus rapides possible, sollicitant un temps de jugement le plus court possible, afin d'attribuer du temps à des tâches davantage valorisées.

Par conséquent, les formes d'évaluation à l'œuvre à l'université sont de nature à opérer un classement des étudiants afin de pouvoir sélectionner ceux qui seront les plus capables de continuer leur formation : c'est l'évaluation « *summative* ». Alors que cette forme de test devrait permettre de valider l'acquisition des compétences spécifiques à une économie de la connaissance, elle provoque une distanciation des étudiants vis-à-vis de la construction d'une motivation intrinsèque pour l'apprentissage, ce qui réduit l'esprit critique de l'étudiant sur son environnement. Malgré ces conséquences négatives, l'évaluation resterait résistante au changement et serait le dispositif le plus conservateur de l'université (Bloxham et al., 2016).

¹ Post-doctorant en économie, ERUDITE (EA 437), Université Paris-Est Créteil — TEPP (FR 3435) — Centre d'étude de l'emploi et du travail — CNAM. remi.le-gall1@gmail.com.

Pourtant, l'examen « *formative* » a également été conceptualisée (Black et William, 1998). Elle intègre le retour sur le résultat, favorise la transparence des règles, intègre l'explicitation précise de l'articulation du critère d'évaluation et les objectifs à atteindre, permet d'augmenter des compétences de jugement des pairs et d'autoévaluation.

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. 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 ?

Nous avons tenté de répondre à cette question en expérimentant une forme nouvelle d'examen 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 d'une licence d'Économie-Gestion. Dans ce cadre, il s'agit de rompre la balance du pouvoir du test où 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 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. Elle permettrait de renforcer des compétences qui conduisent les étudiants à améliorer leurs résultats à l'université.

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. 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.

1. Négocier les règles d'évaluation...

1.1. ... l'apport de la microéconomie de l'éducation et des contrats

En théorie, quand l'éducation est envisagée comme un bien public par les deux pôles de la relation d'instruction, les enseignants et les étudiants, la coopération qui tient compte de l'interdépendance entre les niveaux d'effort d'apprentissage et de l'implication dans l'instruction produit des niveaux d'acquisition de connaissances supérieurs à une situation sans coopération (Correa et Gruver, 1986).

En revanche, lorsqu'il existe des divergences d'intérêt et d'attentes sur ce que procure l'éducation, un problème d'incitation doit être résolu, parce qu'il existe un aléa moral sur l'effort produit par les agents, les étudiants (Michaelis et Schwanebeck, 2016 ; Zubrikas, 2016). Ainsi, le dispositif d'évaluation des connaissances est le levier utilisé par le principal, l'enseignant, qui permet d'aligner les intérêts des deux parties, en conditionnant l'effort d'apprentissage des étudiants à la passation de tests et de l'obtention de notes. En revanche, il peut détériorer la motivation intrinsèque à réaliser l'effort (Frey et Jegen, 2001).

Un moyen de tenir compte de l'ambivalence du dispositif d'évaluation est de proposer de négocier son paramétrage. En effet, une hypothèse traditionnelle de la relation d'agence est que le principal offre un contrat « à prendre ou à laisser », dans lequel il possède tout le pouvoir de négociation. Pourtant, dans des situations organisationnelles traditionnelles, les deux parties détiennent des marges de manœuvre pour négocier les termes du contrat (Demougin et Helm 2004 ; Dittrich et Städter, 2015). Dans ce contexte, y compris en présence d'aléa moral, l'agent peut utiliser ce pouvoir de négociation, pour atteindre un optimum social plus efficace qu'en situation sans compromis.

Dans le cadre de la relation étudiant-enseignant, nous soutenons que la négociation de la règle d'examen dont les termes relèveraient des critères de jugement, de la pondération de ces tests, ou du niveau d'effort attendu de l'étudiant pour acquérir des connaissances, améliorerait la situation des deux individus. Elle contribuerait 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.

1.2. ... l'apport des sciences de l'éducation

Les chercheurs en sciences de l'éducation ont appliqué ces principes de négociation des modalités de l'évaluation des connaissances. Le concept clé de ce champ de recherche est le partenariat dans l'évaluation (Bovill, et al., 2011) dans lequel les rôles changent : 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. 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.

La plupart des essais de partenariats pratiquent la co-construction de la grille d'évaluation (Falchikov, 1986 ; Stefani, 1998 ; Deeley et Bovill, 2017 ; Leslie et Gorman, 2017). Dans ce contexte, l'enseignant et les étudiants négocient les critères d'examen et le niveau d'acquisition des connaissances et compétences. Le plus souvent, l'évaluateur propose une liste de critères, amendés par les étudiants. D'autres études proposent des formules ou des menus de différentes strates d'examens qui sont ensuite soumis au choix des étudiants (Garside et al., 2009 ; Abdelmalak, 2016). Mettre à contribution les étudiants pour créer leur propre mode d'évaluation est également envisageable (Doyle et al., 2017).

Ces interventions provoquent 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, 1998 ; 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 peut passer d'autocrate à celui d'accompagnant disposé à concilier les aspects de transmission de connaissances avec ceux de l'évaluation.

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 » (Deeley et Bovill, 2015 ; Abdelmalak, 2016). Dans la mesure où les grilles sont 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'innovation dans la façon dont est appréciée l'acquisition de connaissances.

Finalement, la plupart de ces travaux ne se focalisent pas explicitement sur des indicateurs de réussite (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. Même si la majorité des études s'accordent sur le fait que les étudiants sont davantage motivés, la question non résolue est celle de savoir si le gain de motivation se traduit par un gain de performance.

2. Le choix de l'approche expérimentale dans un contexte universitaire

En tenant compte des avantages et des limites de ces initiatives précurseuses, 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)². Le taux de réussite de l'Upem se situe ainsi 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 %

² L'Upem a été créée par le décret n°91-706 du 22 juillet 1991 dans le cadre du plan de modernisation « Université 2000 ». Elle a intégré au 1^{er} janvier 2020, l'université cible financée par le plan I-SITE du PIA3, Université Gustave Eiffel (décret n°2019-1360 du 15 décembre 2019).

des étudiants ne se réinscrivent pas au bout de la première année. Entre 2007 et 2016, les taux de passage en deuxième année fluctuent autour de 45 %. C'est donc un environnement propice à l'amélioration des conditions de réussite des étudiants.

2.1. Le protocole expérimental

L'expérimentation s'est déroulée lors du premier semestre de l'année 2017-2018, au sein de plusieurs travaux dirigés (TD) de première année (L1) et de deuxième année (L2) de licence. Nous décrivons la séquence du traitement qui devait permettre de faire émerger une discussion et une négociation entre les étudiants et l'enseignant sur les modalités de l'évaluation de leur participation dans ces TD.

Lors de la première séance, le chargé de TD énonçait les règles de l'expérimentation à chaque groupe test 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. À la fin de cette séance, 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 TD, l'enseignant demandait aux étudiants de donner leurs propres critères. Le pouvoir de décision était donc d'abord dans les mains des étudiants. Ce principe permettait de ne pas créer de jugements préétablis et de ne pas restreindre la créativité ou l'imagination des groupes. À la fin de la discussion, une liste de critères consensuels était arrêtée.

Entre la deuxième et la troisième semaine de TD, l'enseignant retouchait les critères débattus par les classes en supprimant ou ajoutant des critères. Dans cette phase, l'enseignant décidait également de la pondération des listes. Le principe appliqué était de considérer que tous les critères (étudiants ou enseignant) donnaient le même nombre maximal de points. Si un critère se vérifiait durant les deux tiers des séances de TD, il donnait le maximum de points (20) ; s'il se vérifiait durant le tiers des séances de TD, il donnait la moitié de points (10) ; s'il ne se vérifiait jamais, il ne donnait aucun point (0).

Au début de la troisième séance de TD, l'enseignant distribuait la fiche d'évaluation de la participation. 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, chaque groupe test obtenait une grille d'évaluation qui était spécifique à chaque discussion.

À la fin du semestre, la moyenne arithmétique de ces éléments était calculée pour constituer la note de participation pour chaque étudiant, en tenant compte de ces grilles négociées. Ces notes étaient transmises au secrétariat pédagogique pour le calcul de la note globale de contrôle en continu. Lors de la dernière séance, 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. Une appréciation générale de la participation était rédigée afin de donner des recommandations pour la suite du cursus de formation.

2.1. Les participants à l'expérimentation

2.1.1. Modalités d'assignation aléatoire

Trois modules d'introduction à l'économie en L1 et deux modules de microéconomie en L2 ont été choisis pour appliquer ce protocole. Le choix des modules dépendait du service d'enseignement d'un attaché temporaire d'enseignement et de recherche (ATER) et de préoccupations sur la volonté d'améliorer les résultats des étudiants dans ces modules peu favorables à la réussite. Mettre en place le dispositif expérimental dans les groupes d'un seul chargé de TD permettait de contrôler les biais pédagogiques propres à chaque tuteur.

Les groupes « tests » ont été distribués aléatoirement à l'enseignant par le secrétariat pédagogique, en fonction de ses choix d'emploi du temps et de l'organisation générale du planning de formation. Au sein de ces groupes, les étudiants sont répartis selon un panachage aléatoire du secrétariat pédagogique, qui dépend des choix d'options et de l'ordre alphabétique des noms de famille. Ainsi, les étudiants ne « choisissent » pas leur emploi du temps. Aucun changement n'est autorisé en cours de semestre. Les groupes « témoins » sont constitués selon les mêmes principes que les groupes tests. De cette manière, on se rapproche d'une expérimentation contrôlée, au même titre par exemple que dans le cadre de l'expérimentation du Projet Voltaire (Bellity et al., 2017).

Afin de se prémunir contre des biais d'organisation liés à la condition expérimentale (L'Horty et Petit, 2011), un comité de suivi composé des responsables de chaque niveau de formation a encadré les étapes du protocole et les enseignants-référents des modules étaient informés du dispositif. Nous avons également demandé le consentement des étudiants d'utiliser leurs informations personnelles : 70 % des étudiants des deux promotions ont accepté l'utilisation encadrée de leurs données personnelles.

2.1.2. Échantillon

Lors du premier semestre de l'année 2017-2018, 289 étudiants étaient inscrits en L1 et 202 étudiants étaient inscrits en L2. Nous avons écarté de cette population les étudiants en contrôle terminal ainsi que ceux n'ayant pas consenti à l'utilisation de leurs résultats aux examens.

En L1 (tableau 1), il y a davantage d'étudiants 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. 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 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.

En ce qui concerne la promotion de L2 (tableau 2), nous observons également davantage d'étudiants 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. 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. Il y a significativement moins de redoublants dans le groupe traité.

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)	13,430** (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 TD	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.

2.2. La stratégie d'identification des effets

Les données utilisées pour estimer les effets de la négociation de l'évaluation de la participation en TD sur la réussite proviennent de trois sources :

- de l'application pour l'organisation et la gestion des enseignements et des étudiants (Apogée) qui contient pour chaque module, les notes finales du contrôle continu et du contrôle terminal ;
- des « Fichiers de Gestion des Absences et des Résultats/notes de TD (Figaro) » conçus par les membres de l'UFR et qui permettent de suivre l'assiduité, d'identifier les groupes de travaux dirigés et la décomposition des notes du contrôle continu, ainsi que leur pondération pour chaque module ;
- de la plateforme de saisie des vœux d'inscription dans l'enseignement supérieur Admission Post-Bac (remplacée au printemps 2018 par la plateforme Parcoursup), qui donne les

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, donc hors réorientations internes à l'université et hors bacheliers ayant été affectés par le rectorat pendant la phase complémentaire d'affectation.

Nous avons mené deux stratégies d'identification afin de contrôler l'hétérogénéité de la difficulté de toutes les composantes des formations qu'elles aient été altérées ou non par l'expérimentation. D'une part, nous considérons les trois ensemble de notes semestrielles des modules dans lesquels est intervenu le processus expérimental et des modules qui n'ont pas modifié leur environnement d'évaluation : lors des deux examens en contrôle continu ; lors de l'examen partiel en contrôle terminal. Ainsi, on peut 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 unités d'enseignement 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 obtenu de meilleurs résultats dans le module traité par rapport au groupe témoin et leurs résultats dans les autres modules. Nous pouvons ainsi attribuer à l'environnement de négociation un effet causal sur la réussite. Les autres paramètres du modèle mesurent d'autres indicateurs d'intérêt :

- $\beta = \overline{Y_{00}}$, correspond à la note moyenne obtenue par le groupe témoin dans les modules non traités ;
- $\alpha_1 = \overline{Y_{10}} - \overline{Y_{00}}$, identifie l'écart de notes moyennes entre le groupe traité et le groupe témoin pour les modules de formation non traités ;
- $\alpha_2 = \overline{Y_{01}} - \overline{Y_{00}}$, 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.

D'autre part, nous prenons en compte le fait d'avoir validé ou non le semestre lors de la première session d'examens. 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 lors du semestre. Dans ce cadre, les notes seules ne jouent pas sur l'admission, la pondération des modules et les règles de compensation entrent aussi en considération dans l'effort d'apprentissage. Nous recourons à un modèle dichotomique de type logit afin de modéliser la probabilité de valider le semestre. 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. 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.

3. Des résultats dépendant de l'exécution du contrat d'évaluation négocié

3.1. En L1, une perméabilité plus facile à l'intérêt de la négociation

Les résultats de l'expérimentation sur la réussite aux examens des étudiants de L1 sont inscrits dans le tableau 3. Nous constatons un effet positif et significatif de la condition expérimentale, à la fois lors du premier examen en contrôle continu et lors du deuxième examen en contrôle continu (colonne 1 et colonne 3). Lorsqu'on tient compte des matières qui n'ont pas intégré la négociation de la règle d'évaluation de la participation, l'effet positif se maintient (colonne 2 et 4). Enfin, les étudiants bénéficiant du dispositif expérimental ont mieux réussi dans le module traité que dans les autres modules. Cet écart de réussite est deux fois plus important lors du second examen que lors du premier examen en contrôle continu et il est significatif.

Néanmoins, ces constats positifs disparaissent 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.

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

Pour finir, nous constatons un effet positif, mais non significatif de la condition expérimentale sur la probabilité de valider le semestre pour les L1 y compris en incluant des variables de contrôle (tableau 4). 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.

Tableau 4 : 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)
Référence : Néo-entrants					
Redoublants				-0,056 (0,165)	-0,023 (0,156)
Réorientation				-0,026 (0,112)	-0,053 (0,132)
Référence : Baccalauréat E.S					
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
Critère AIC	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. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

3.2. En L2, une méfiance initiale compensée par l'application du contrat

L'évaluation de l'impact de la condition expérimentale sur les scores des étudiants de L2 conduit à des résultats opposés à ceux de première année (tableau 5). En effet, le groupe traité a moins bien réussi que le groupe témoin aux deux examens en contrôle continu. Ce résultat est plus significatif lors du deuxième contrôle des connaissances (colonne 3). L'écart négatif des notes 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 notes 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. Par conséquent, l'adoption d'une pratique de négociation aurait 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 (colonne 5). Ces groupes 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.

Tableau 5 : 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. * $p < 0,10$, ** $p < 0,05$, *** $p < 0,01$.

Malgré ces constats pessimistes, 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 (tableau 6). 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 auraient adopté des stratégies d'optimisation de leurs résultats dans les autres matières.

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
Critère AIC	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.

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

3.3. L'analyse de l'hétérogénéité des résultats

Nous obtenons des résultats hétérogènes et des effets qui semblent opposés : un gain de performance pour les L1 ; une détérioration de performance pour les L2. 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.

Nous montrons ainsi que lorsque le contexte n'est pas favorable à la négociation, 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.

4. Une approche à concrétiser et à développer dans d'autres organisations

En conclusion, nous contribuons à l'ensemble des études de mesure d'impact de dispositifs de remédiation dans l'enseignement supérieur, en développant l'idée selon laquelle l'un des déterminants de l'engagement des étudiants dans leur apprentissage se noue dans la relation qu'ils créent avec l'équipe enseignante. 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).

Nous montrons qu'il est possible d'expérimenter des outils innovants pour modifier 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. Une meilleure technique 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.

Sous certaines conditions, nous apportons des arguments favorables à la mise en place d'environnements de négociation des dispositifs de vérification de l'information. Dans des relations d'agence, 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.

Cependant, les caractéristiques de l'expérimentation soulèvent un problème de validité externe : la taille des effectifs est faible en comparaison d'études expérimentales sur la réussite (Dmitrijeva et al., 2014) ; le fait qu'elle ait été réalisée au sein d'une unique unité de formation et de recherche, dans une seule université ; les partenariats évaluatifs ont été menés avec un seul enseignant. 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.

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. Ces deux formes d'évaluation revêtent les enjeux 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. 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 et ne pas produire de situations préjudiciables pour les étudiants.

Bibliographie

- Abdelmalak, M. M. M. (2016). Faculty-Student Partnerships in Assessment. *International Journal of Teaching and Learning in Higher Education*, 28 (2), p. 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.
- 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). Travail, emploi et politiques publiques.
- Black, P., & William, D. (1998). Assessment and Classroom Learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), p. 7-74.
- 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), p. 466-481.
- Boud, D., & Falchikov, N. (Éds.). (2007). *Rethinking assessment in higher education: Learning for the longer term*. Routledge.

- 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), p. 133-145.
- Correa, H., & Gruver, G. W. (1987). Teacher-student interaction: A game theoretic extension of the economic theory of education. *Mathematical Social Sciences*, 13(1), p. 19-47.
- Deeley, S. J., & Bovill, C. (2017). Staff student partnership in assessment: Enhancing assessment literacy through democratic practices. *Assessment & Evaluation in Higher Education*, 42(3), p. 463-477.
- Deeley, S. J., & Brown, R. A. (2014). Learning Through Partnership in Assessment. *Teaching and Learning Together in Higher Education*, 12, 10 p.
- Demougín, D., & Helm, C. (2006). Moral Hazard and Bargaining Power. *German Economic Review*, 7(4), p. 463-470.
- Dittrich, M., & Städter, S. (2015). Moral hazard and bargaining over incentive contracts. *Research in Economics*, 69(1), p. 75-85.
- Dmitrijeva, J., Parquet, L. du, L'Horty, Y., & Petit, P. (2014). Réussir à l'université avec un emploi salarié. *Revue française d'économie*, XXIX (1), p. 155-187.
- 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*, 24(6), p. 1-16.
- Falchikov, N. (1986). Product comparisons and process benefits of collaborative peer group and self-assessments. *Assessment & Evaluation in Higher Education*, 11(2), p. 146-166.
- Frey, B. S., & Jegen, R. (2001). Motivation Crowding Theory. *Journal of Economic Surveys*, 15(5), p. 589-611.
- 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), p. 141-148.
- Hölmstrom, B. (1979). Moral Hazard and Observability. *The Bell Journal of Economics*, 10(1), p. 74-91.
- 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, p. 220-225.
- L'Horty, Y., & Petit, P. (2011). Évaluation aléatoire et expérimentations sociales. *Revue française d'économie*, XXVI(1), p. 13-48.
- Le Gall, R. (2018). L'économie des dispositifs de vérification de l'information : une approche expérimentale. Thèse pour l'obtention de doctorat en économie. Université Paris-Est.
- 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), p. 286-301.
- McMillan, J. H. (Éd.). (2013). *Handbook of research on classroom assessment*. Sage Publications.
- Michaelis, J., & Schwanebeck, B. (2016). Examination rules and student effort. *Economics Letters*, 145, p. 65-68.
- 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), p.152-167.
- Morlaix, S., & Perret, C. (2013). L'évaluation du Plan Réussite en Licence : quelles actions pour quels effets ? *Recherches en Éducation*, (15), p. 137-150.
- Razafindratsima N., & Bonnevalle L. (2019), Parcours et réussite en licence : les résultats de la session 2018, Note Flash n°26, MESRI — SIES.
- Stefani, L. A. J. (1998). Assessment in Partnership with Learners. *Assessment & Evaluation in Higher Education*, 23(4), 339-350.
- Zubrickas, R. (2015). Optimal grading. *International Economic Review*, 56 (3), p. 751-776.