



**HAL**  
open science

# Conjuguer des méthodes quantitatives et qualitatives pour évaluer un dispositif d'innovation pédagogique à l'Université

Laurent Gensbittel, Muriel Epstein

## ► To cite this version:

Laurent Gensbittel, Muriel Epstein. Conjuguer des méthodes quantitatives et qualitatives pour évaluer un dispositif d'innovation pédagogique à l'Université. 33e colloque de l'ADMEE-EUROPE: "Pluralité de contextes, pluralité d'évaluations en éducation: quelles interactions et quels enjeux?", ADMEE-EUROPE, Apr 2022, Pointe-à-Pitre, France. hal-03906865

**HAL Id: hal-03906865**

**<https://hal.science/hal-03906865>**

Submitted on 19 Dec 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Conjuguer des méthodes quantitatives et qualitatives pour évaluer un dispositif d'innovation pédagogique à l'Université.

Laurent Gensbittel, ingénieur pédagogique à l'université Paris 1 Panthéon Sorbonne et doctorant au laboratoire EMA, Université CY Cergy Paris, [laurent.gensbittel@etu.cyu.fr](mailto:laurent.gensbittel@etu.cyu.fr)

Muriel Epstein, MCF, laboratoire EMA Université CY Cergy Paris, [muriel.epstein@cyu.fr](mailto:muriel.epstein@cyu.fr)

**Axe du colloque:** Axe 3 : L'évaluation de dispositifs et de politiques de formation : comment conjuguer des sources multiples ?

**Résumé court:** (250 à 300 mots)

Accompagner des pratiques d'innovation en éducation par la recherche permet d'inscrire les enseignants dans une dynamique de recherche où ils évaluent leur propre pratique (Rege Colet et al., 2011).

Lors de cette recherche, une équipe pluridisciplinaire a comparé les effets de deux formes de travaux dirigés (TD) en troisième année de licence d'économie: une classe a réalisé de courtes vidéos sur des mécanismes clés en macroéconomie, tandis que les autres classes recevaient les TD sous format usuel. Le cours d'amphithéâtre était le même pour tous les étudiants, ainsi que l'examen final. L'expérience a été étendue, mais modifiée, l'année suivante à cinq classes tests.

Cinq étudiants de la première vague ont été interrogés lors d'entretiens avant et après l'examen final. Les étudiants des deux vagues ont également répondu à un questionnaire en ligne. Les enseignants ont participé à des entretiens cliniques analysés selon la méthode de la théorie enracinée (Glaser & Strauss, 2010). Enfin, les données statistiques anonymisées des situations sociologiques et des résultats aux examens de Licence 3 concernant la totalité des étudiants ont été étudiées.

Les modèles quantitatifs ne concordent pas avec les entretiens qualitatifs. Les premiers montrent de bons résultats la première année bien que les étudiants "expérimentaux" ne se sentent pas préparés. La seconde année, c'est surtout un effet sur le développement professionnel des chargés de TD qui est perceptible dans les entretiens sans que cela ne se traduise quantitativement dans les notes des étudiants.

L'objectif de cette communication est de présenter l'articulation entre des méthodes quantitatives et qualitatives pour évaluer un dispositif d'innovation pédagogique. De fait, l'évaluation de l'expérience dépend de ce que l'on considère tant pour les étudiants (ressenti, résultats aux examens) que pour les chargés de TD (formation, sentiment de compétence, résultats réels pour les étudiants à l'examen).

**Mots clés:** pédagogie du supérieur, méthodes, numérique, vidéo, économie

## Introduction

En 2018-2019, une équipe composée de deux enseignants d'économie, d'une chercheuse en sociologie de l'éducation et d'un ingénieur pédagogique a conduit une classe de travaux dirigés de 31 étudiants en licence d'économie à l'université Paris 1 Panthéon Sorbonne à réaliser de courtes vidéos (4-5 minutes) sur des mécanismes clés en Macroéconomie monétaire internationale. Le groupe de TD a été tiré au sort parmi 2 groupes encadrés par le même

chargé de TD qui étaient tous deux indiqués comme “TD expérimentaux”. Le second TD, qui a suivi un parcours classique mais observé, a servi de groupe contrôle. Par ailleurs, 21 autres groupes (556 étudiants) recevaient les TD sous leur format usuel. Le cours d’amphithéâtre quant à lui était le même pour tous les étudiants, ainsi que l’examen final.

Pour l’enseignante porteuse de projet, il y avait plusieurs objectifs, outre les objectifs pédagogiques liés à la matière et communs à tous les étudiants, portant globalement sur la maîtrise des principaux mécanismes de macroéconomie. D’abord il s’agissait de rendre les séances de TD plus efficaces en termes d’apprentissage que les TD classiques et donc de savoir si faire travailler des étudiants en groupes sur la production de vidéo, chaque groupe sur un sujet précis, pendant plusieurs séances se verrait sur les copies de l’examen final commun à tous les groupes de TD. Ensuite, il importait de pouvoir travailler des compétences transversales comme la créativité et le travail en équipe. Enfin, elle souhaitait, par l’expérience, former et fidéliser les chargés de TD, soit une dizaine de vacataires et doctorants. L’expérience a été renouvelée en 2019-2020 et transformée. En effet, elle a été élargie à 4 chargés de TD. Ainsi 10 groupes appariés (ayant le même chargé de TD et volontaires pour l’expérience) ont participé dont 5 groupes test (158 étudiants) et 5 groupes contrôles (150 étudiants), 6 groupes (220 étudiants) étaient eux en format usuel et hors expérimentation. Le contexte a cependant largement modifié les conditions d’expérimentation : diminution du nombre d’heures de TD passées de 12h à 10h, alourdissement du travail des étudiants en TD expérimentaux, contexte de grève des transports qui a retardé l’examen de 2 mois. Enfin, le cours est passé du dernier semestre de licence 3 au premier.

Cette communication examine les articulations entre les méthodes d’évaluation qualitatives et quantitatives d’une expérience qui visait simultanément l’amélioration des résultats pour les étudiants à l’examen et la professionnalisation des chargés de TD.

## Cadre théorique

Accompagner des pratiques d’innovation en éducation par la recherche permet d’inscrire les enseignants dans une dynamique de recherche où ils évaluent leur propre pratique et progressent (Rege Colet et al., 2011). Le fait d’intégrer une évaluation par les étudiants de leurs cours favorise le développement professionnel des enseignants (Berthiaume et al., 2011; Younes, 2020). Reste qu’évaluer des dispositifs innovants nécessite une réflexion encore à conduire sur les cadres de références et les objectifs (Buard, 2020; Burton et al., 2013).

Bucheton et Soulé (2009) distinguent deux étayages de la part des enseignants : les postures de “contrôle” où l’enseignant est très guidant des postures de “lâcher prise” ou d’accompagnement dans lequel l’enseignant s’adapte à l’avancée des élèves ce qui favorise leur engagement.

Bandura (2006) montre que le sentiment d’efficacité personnelle s’articule avec le comportement et l’environnement. Lorsque ces trois éléments sont alignés, alors les croyances renforcent le sentiment d’efficacité personnelle, dans le cas inverse, la personne peut adopter des comportements inadaptés.

## Méthodologie

La première année, les TD tests et contrôle ont été observés lors de 5 séances chacune par les auteurs. Huit étudiants du TD expérimental ou du TD de contrôle ont été interrogés lors d’entretiens individuels ou collectifs par la sociologue avant ou après l’examen final. Les étudiants ont également répondu à un questionnaire en ligne (n=174 dont n=59 pour les groupes expérimentaux et contrôles). Les deux enseignants d’économie ont répondu à des entretiens individuels non directifs, analysés selon la méthode de la théorie enracinée (Glaser & Strauss, 2010). Enfin, les données statistiques anonymisées sociodémographiques et les résultats aux examens de L3 concernant la totalité des étudiants (n=647) ont été étudiées.

La seconde année, ont été administrés un questionnaire à 2 ans pour les étudiants de la première vague (n=24) et un questionnaire en ligne aux étudiants de la seconde vague similaire à l'année précédente (n=136 dont 118 des TD expérimentaux et contrôle). Les grèves de transport ont fait diminuer le taux de participation et n'ont permis qu'une seule observation de classe. Cependant des TD filmés destinés à compenser le nombre d'heures de TD ont permis la visualisation de TD. Des entretiens non directifs de deux heures avec trois des quatre chargés de TD expérimentaux ont été conduits, des entretiens ont été menés avec deux étudiantes du quatrième chargé de TD. La modélisation quantitative avec les données sociodémographiques a porté sur 332 participants et non la totalité des 528 étudiants afin de disposer des notes du semestre précédent du fait du décalage du cours au premier semestre. L'année précédente, il avait été possible de récupérer les informations du premier semestre de Licence 3 pour tous les étudiants.

Les statistiques ont été faites avec des modèles de régression et des modèles ANCOVA pour l'essentiel afin de d'identifier l'existence d'un effet de différents facteurs sur les résultats des étudiants.

## Résultats

### *Résultats académiques et apprentissages perçus par les étudiants*

La première année, les résultats des étudiants de la classe expérimentale à l'examen final s'avèrent significativement supérieurs de 1,6 points (sur 20) à tous les autres étudiants de la promotion, une fois pris en compte les autres facteurs de réussite tels que les résultats obtenus aux examens du semestre précédent dans un modèle de régression linéaire. De manière intéressante, l'effet positif touche aussi bien les meilleurs étudiants que la queue de classe. La seconde année, les résultats quantitatifs sont peu significatifs (à part l'effet Hawthorne qui apparaît clairement).

Pourtant les données qualitatives issues des entretiens (« je pense que je serai moins préparé pour le partiel car ça reste un partiel classique » Entretien collectif du 21 février 2019) et les données quantitatives issues des questionnaires pointent un problème de conscientisation des apprentissages qui peut apparaître avec les pédagogies actives ((Deslauriers et al., 2019)). A la question « A quel point le TD vous a-t-il aidé dans vos apprentissages ? », notée de 0 à 10, la moyenne est à 6,01 pour le TD expérimental et à 6,75 pour les autres TD la première année, à 5,83 pour le TD expérimental et à 8,15 pour les groupes contrôles la seconde année. Ainsi les étudiants des groupes expérimentaux se sentent significativement moins bien préparés que les autres, alors que, comme indiqué plus haut, la première année, leurs résultats sont significativement<sup>1</sup> meilleurs à l'examen de fin d'année, la seconde année, leurs résultats sont équivalents. On note donc une disjonction entre l'évaluation qualitative et quantitative.

### *Charge de travail des étudiants*

On peut faire une observation similaire sur la différence entre les perceptions et les chiffres concernant le temps de travail. Dans le questionnaire, plusieurs commentaires libres d'étudiants de TD expérimentaux pointent les aspects chronophages de la réalisation de vidéo (« l'idée de la vidéo est intéressante mais très chronophage », « Bonne expérience mais double de travail à effectuer », étudiants de la seconde année). Mais dans le même questionnaire, sur l'échelle d'effort ("Quel effort avez-vous mis dans le TD ?") et sur le temps de travail personnel, les étudiants de TD expérimentaux indiquent un effort et un temps de travail identiques voire inférieurs à ceux des autres étudiants.

---

<sup>1</sup> Les significativités sans autre précision correspondent systématiquement à des tests de Students avec une p-value <0.05

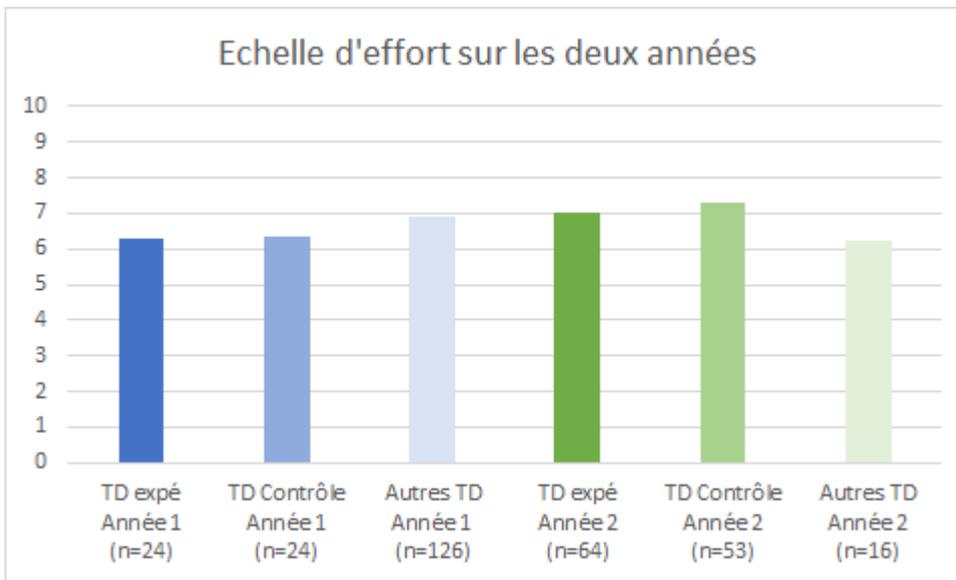


Figure 1 : Réponses au questionnaire des deux années à la question "Quel effort avez-vous mis dans le TD ?"

L'effort indiqué par les étudiants du TD expérimental est le plus faible la première année et est plus faible la seconde année que celui indiqué par les étudiants des TD contrôles.

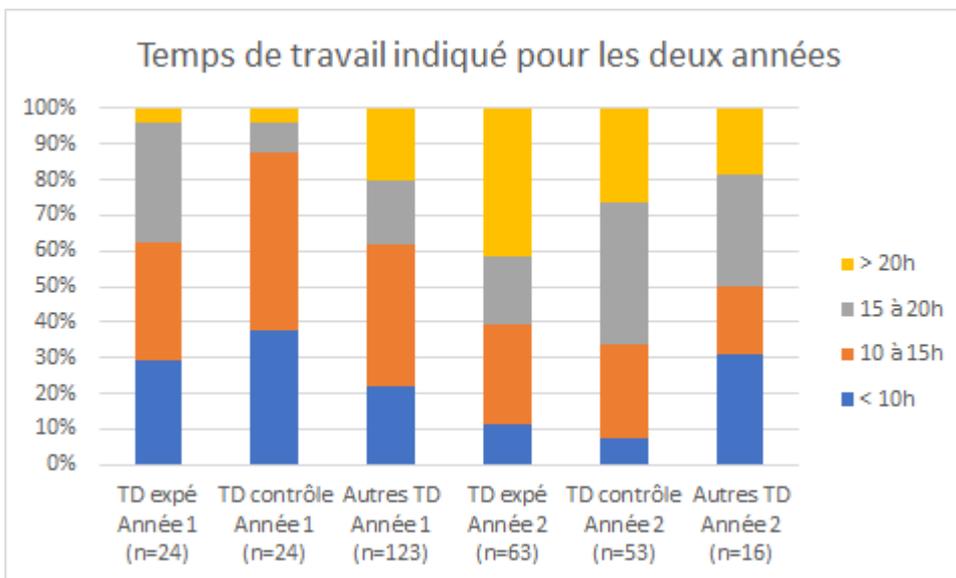


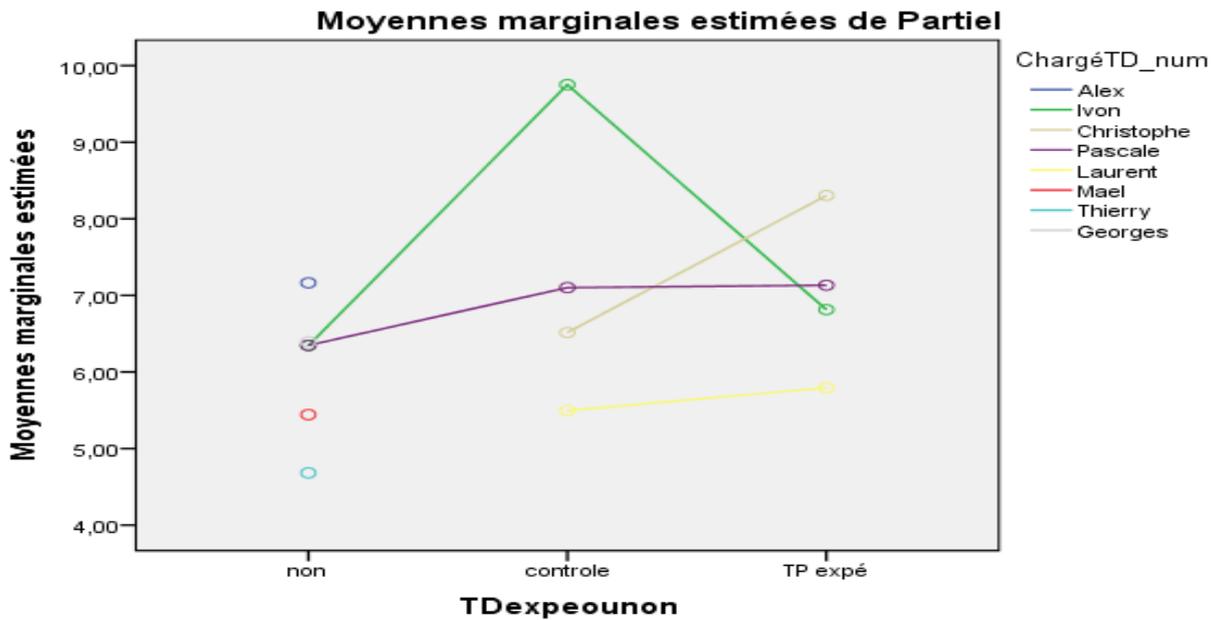
Figure 2 : Réponses au questionnaire des deux années à la question "Combien de temps personnel, en dehors des heures de TD, avez-vous consacré au TD ?"

Le temps de travail personnel indiqué par les étudiants du TD expérimental est plus faible la première année que celui des autres TD. La seconde année, les étudiants des TD expérimentaux ont été plus nombreux à indiquer avoir travaillé plus de 20h, mais un peu moins nombreux que les étudiants des TD contrôles à avoir travaillé plus de 15h. Nous pouvons aussi constater une dissociation entre l'effort et le temps de travail la première année en comparant les indications des étudiants des TD expérimentaux et de contrôle, comme si les étudiants des TD expérimentaux affirmaient travailler davantage en ressentant moins l'effort. C'est sans doute à rapprocher de ce qui ressort de certains entretiens indiquant une motivation et une entraide accrue dans les groupes expérimentaux et le sentiment d'apprendre "sans chercher à apprendre" (étudiante 1 - juin 2019).

### Effets des chargés de TD

La première année, le chargé de TD a été globalement très apprécié par les étudiants, certains affirmant même que la réussite de l'expérimentation lui devait beaucoup. La seconde année, il n'a pris en charge aucun TD mais a formé les chargés de TD impliqués dans l'expérimentation et ceux-ci pouvaient le solliciter en cas de besoin. Nous avons donc

voulu voir si les résultats au partiel étaient différents d'un chargé de TD à l'autre et s'il pouvait y avoir un « effet des chargés de TD ».



Les covariables apparaissant dans le modèle sont évaluées pour les valeurs suivantes : semestre4 = 11,45921

Les moyennes non estimées ne sont pas représentées

Figure 3 : Résultats au partiel la 2<sup>de</sup> année en fonction des chargés de TD

Nous constatons, à travers un modèle ANCOVA, que les résultats au partiel ont été les meilleurs pour les étudiants du TD témoin animé par Ivon<sup>2</sup> avec une moyenne proche de 10. Les deux autres TD d'Ivon (TD expérimental et TD hors expérimentation) ont des résultats peu différents avec une moyenne proche de 6,5. La différence de moyenne entre TD expérimental et TD témoin n'est pas nette pour les groupes de Laurent et de Pascale. Par contre, les résultats sont meilleurs pour les étudiants du TD expérimental de Christophe que pour ceux de son TD témoin.

Nous avons essayé de formaliser les postures des chargés de TD, telles que nous les avons perçues et/ou qu'ils ont pu nous les décrire au cours d'entretiens. Certains d'entre eux ont adopté une posture stricte, directive, pendant les TD expérimentaux à l'inverse d'autres adoptant une posture souple, plus adaptative et réactive. De même, certains étaient à l'aise avec la matière et l'expérimentation, sentiment que l'on peut rapprocher du sentiment d'efficacité personnelle (Bandura, 2006) tandis que d'autres n'étaient pas à l'aise pour diverses raisons, voire adoptait une attitude fuyante (TD peu ou pas préparés, plusieurs absences), dans le cas de Christophe.

### Le sentiment d'efficacité et la posture des chargés de TD

Chargé.es de TD de la 2 <sup>de</sup> année	Sentiment d'efficacité	Posture souple (accompagnante)
Ivon	+	-
Pascale (comme Brendan la 1 <sup>ère</sup> année)	+	+
Laurent	-	-
Christophe	-	+

<sup>2</sup> Alors que Brendan Harnoy-Vannier, chargé de TD de l'année 1 a été coauteur et est donc nommé sous son vrai prénom, les chargés de TD de l'année 2 sont anonymisés

L'articulation du tableau 1 avec le graphique Figure 3 tend à indiquer, d'une part, que les chargés de TD avec un fort sentiment d'efficacité aident leurs étudiants à surperformer là où les chargés de TD mal à l'aise, avec un faible sentiment d'efficacité, les font sous-performer. D'autre part, l'expérience a un effet positif dans le cas où les chargés de TD sont souples tandis qu'elle n'en a pas spécialement s'ils sont stricts, autrement dit qu'il y a une adéquation importante entre le style pédagogique du chargé de TD (strict ou souple) et l'expérience (en pédagogie active).

Notre hypothèse est que dans le cas d'un chargé de TD défaillant, Christophe, les étudiants ont vite compris qu'ils ne pouvaient compter que sur leur travail personnel. Ainsi, ses étudiants en TD expérimentaux ont eu accès à une méthode de travail de groupe en autonomie qui leur a permis de compenser son manque d'investissement, ce que n'ont pas eu ses autres étudiants. À l'inverse, les postures très guidantes de Laurent, dont l'ensemble des étudiants a sous-performé, et d'Ivon, qui a fait surperformer les siens, n'a sans doute pas permis aux étudiants de leurs TD expérimentaux de profiter de la dynamique du travail de groupe et de s'engager pleinement dans l'expérimentation.

### Effets sur les chargés de TD

Concernant les aspects formateurs voire transformateurs de l'expérimentation sur les chargés de TD, on constate un changement de positionnement par et sur l'innovation pédagogique (« j'aurai plus confiance pour expérimenter [d'autres] formats » Ivon la 2<sup>nd</sup>e année - « Pour moi, la vidéo c'est juste [...] un prétexte [...] pour changer la pédagogie dans le TD. [...] Si j'avais vraiment que moi, je pense que je ferais que des vidéos » Pascale, 2<sup>nd</sup>e année). Nous notons aussi ce changement de rapport au collectif lié à la dynamique de groupe et aux étudiants apprenant les uns des autres (« c'est le côté humain, qu'il y a eu avec l'expérience et le côté collectif de l'expérience. Humain et collectif : ça revient à ce que moi je disais sur la relation prof-élève, où je pense que c'est important de mettre de l'humain là-dedans mais ça revient aussi sur le côté de, apprendre des autres » Brendan, 1<sup>e</sup> année).

Ainsi, cette expérimentation a donné à l'équipe enseignante (professeur et chargé de TD) le sentiment de pouvoir progresser et l'a conduite à réévaluer ses pratiques pédagogiques.

## Conclusion

Nous avons tenté dans cette étude d'articuler différentes méthodes de recherche pour comparer leurs résultats. Les modèles quantitatifs (régression) ne concordent pas avec les entretiens qualitatifs (type cliniques). Les premiers montrent de très bons résultats la première année là où les étudiants ne se sentent pas préparés et ont l'impression, en partie fautive, de travailler plus que les autres, avec un effort perçu moindre. La seconde année, c'est surtout l'effet sur la formation des chargés de TD qui est perceptible dans les entretiens, sans que cela ne semble aider ou pénaliser les étudiants sur les modèles quantitatifs.

Les modèles ANCOVA sur le contrôle d'un effet "chargé de TD" donnent des résultats variables qui, confrontés à l'analyse des entretiens avec les chargés de TD, montrent pourtant que la participation à une expérience est, en soi, formatrice pour eux, avec des effets de renforcement ou de déstabilisation de leurs pratiques pédagogiques selon des critères nécessairement qualitatifs. Le sentiment d'efficacité personnelle du chargé de TD semble avoir un effet positif sur l'ensemble des étudiants, tandis qu'une posture pédagogique souple semble renforcer les bénéfices de méthodes actives.

L'articulation entre les entretiens cliniques sur la vision des chargés de TD de leurs métiers et les résultats quantitatifs aux examens s'avère complexe. De fait, l'efficacité de l'expérience dépend de ce que l'on considère tant pour les étudiants (le ressenti, les résultats aux examens) que pour les chargés de TD (leur formation, leur sentiment de compétence, les résultats réels pour les étudiants à l'examen.) Le ressenti qualitatif des étudiants ainsi que leurs résultats aux examens amènent les enseignants à appréhender différemment leur conception pédagogique.

## Bibliographie

- Bandura, A. (2006). Guide for Constructing Self-Efficacy Scales (Revised). In F. Pajares & T. Urdan (Éds.), *Self-efficacy beliefs of adolescents* (Vol. 5, p. 307-337). Greenwich CT : Information Age Publishing.
- Berthiaume, D., Lanarès, J., Jacqmot, C., Winer, L., & Rochat, J.-M. (2011). L'évaluation des enseignements par les étudiants (EEE). *Recherche et formation*, 67, Art. 67. <https://doi.org/10.4000/rechercheformation.1387>
- Buard, K. (2020). Quelle auto-évaluation de la qualité des dispositifs éducatifs dans les établissements scolaires du second degré ? Proposition d'un nouveau cadre d'analyse. *Évaluer. Journal international de recherche en éducation et formation*, 6(3), 75-104.
- Bucheton, D., & Soulé, Y. (2009). Les gestes professionnels et le jeu des postures de l'enseignant dans la classe : Un multi-agenda de préoccupations enchâssées. *Éducation et didactique*, 3-3, Art. 3-3. <https://doi.org/10.4000/educationdidactique.543>
- Burton, R., Blais, J.-G., & Gilles, J.-L. (2013). TECHNOLOGIES ET ÉVALUATION DANS L'ENSEIGNEMENT SUPÉRIEUR. In M. Romainville, R. Goasdoué, & M. Vantourout, *Évaluation et enseignement supérieur* (1re édition, p. 85-106). De Boeck Supérieur.
- Deslauriers, L., McCarty, L. S., Miller, K., Callaghan, K., & Kestin, G. (2019). Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom. *Proceedings of the National Academy of Sciences*, 201821936. <https://doi.org/10.1073/pnas.1821936116>
- Glaser, B., & Strauss, A. (2010). *La découverte de la théorie ancrée. Stratégies pour la recherche qualitative*. <https://www.decitre.fr/livres/la-decouverte-de-la-theorie-ancree-9782200351175.html>
- Rege Colet, N., McAlpine, L., Fanghanel, J., & Weston, C. (2011). Le concept de Scholarship of Teaching and Learning. La recherche sur l'enseignement supérieur et la formalisation des pratiques enseignantes. *Recherche et formation*, 67, 91-104. <https://doi.org/10.4000/rechercheformation.1412>
- Younes, N. (2020). *Vers une évaluation écologique dans l'enseignement supérieur : Dispositifs et transformations en jeu* [Habilitation à diriger des recherches, Université de Lorraine]. <https://hal.archives-ouvertes.fr/tel-03032417>