



HAL
open science

Continuité et rupture dans les modalités d'évaluation : une activité d'évaluation par les pairs associée à une auto-évaluation au sein d'un cours de didactique des langues

Guilène Révauger

► **To cite this version:**

Guilène Révauger. Continuité et rupture dans les modalités d'évaluation : une activité d'évaluation par les pairs associée à une auto-évaluation au sein d'un cours de didactique des langues. Recherche et pratiques pédagogiques en langues de spécialité - Cahiers de l'APLIUT, 2022, 41 (2), 10.4000/apliut.10239 . hal-03890000

HAL Id: hal-03890000

<https://hal.univ-reunion.fr/hal-03890000>

Submitted on 15 Feb 2023

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.

Copyright

Continuité et rupture dans les modalités d'évaluation : une activité d'évaluation par les pairs associée à une auto-évaluation au sein d'un cours de didactique des langues

*Continuity and Change in Assessments Methods: a Peer Assessment Activity
Combined with Self-assessment in a Language Teaching Course*

Guilène Révauger



Édition électronique

URL : <https://journals.openedition.org/apliut/10239>

DOI : [10.4000/apliut.10239](https://doi.org/10.4000/apliut.10239)

ISSN : 2119-5242

Éditeur

APLIUT

Ce document vous est offert par Bibliothèques de l'Université de La Réunion



Référence électronique

Guilène Révauger, « Continuité et rupture dans les modalités d'évaluation : une activité d'évaluation par les pairs associée à une auto-évaluation au sein d'un cours de didactique des langues », *Recherche et pratiques pédagogiques en langues de spécialité* [En ligne], Vol. 41 N°2 | 2022, mis en ligne le 15 novembre 2022, consulté le 15 février 2023. URL : <http://journals.openedition.org/apliut/10239> ; DOI : <https://doi.org/10.4000/apliut.10239>

Ce document a été généré automatiquement le 8 décembre 2022.

Tous droits réservés

Continuité et rupture dans les modalités d'évaluation : une activité d'évaluation par les pairs associée à une auto-évaluation au sein d'un cours de didactique des langues

Continuity and Change in Assessments Methods: a Peer Assessment Activity Combined with Self-assessment in a Language Teaching Course

Guilène Révauger

Introduction

- 1 La présente note de pédagogie universitaire s'intéresse à une évaluation par les pairs, un dispositif évaluatif relativement récent au sein de notre parcours de formation de master MEEF second degré, langues vivantes. Néanmoins, la réflexion sur l'évaluation entre pairs et ses bénéfiques a débuté il y a plus de cinquante ans, en témoignent les travaux de Burke¹ dès 1969. Le caractère inédit de cette expérimentation est ainsi d'une grande relativité.
- 2 La crise sanitaire puis la réorganisation des maquettes des masters MEEF à la rentrée 2021 ont conduit à la nouvelle scénarisation du cours « Enjeux didactiques de l'enseignement des langues », formation hybride synchrone et asynchrone de l'Inspé de La Réunion. Remodelée, la formation a requis de nouveaux dispositifs d'évaluation faisant appel à un large éventail d'outils. Parmi ces dispositifs figure une évaluation par les pairs associée à une auto-évaluation. Nous en proposons ici l'analyse. Après avoir présenté le cadre de l'étude, nous partagerons les résultats obtenus puis, afin de permettre à d'autres de tirer profit de ce retour d'expérience, nous avancerons un certain nombre de recommandations pratiques.

1. Une évaluation par les pairs

1.1. Définition

- 3 Dans le contexte de notre étude, l'évaluation par les pairs (désormais EP) consiste à solliciter les étudiants afin qu'ils émettent un avis et procèdent à la notation des travaux produits par d'autres étudiants. Si l'on associe l'EP en premier lieu à des modalités d'évaluation utilisées dans les universités américaines (*peer assessment*), l'arbitrage par les pairs est largement répandu en France dans le domaine de la recherche (Condette ; Picard), en témoignent l'évaluation collégiale par les pairs adoptée par l'Agence d'évaluation de la recherche et de l'enseignement supérieur (AERES) puis le Haut Conseil de l'évaluation de la recherche et de l'enseignement supérieur (HCERES) (Glaudes), le recrutement des enseignants-chercheurs par des comités de sélection, ou encore les évaluations de type *peer-reviewing* en double-aveugle auxquelles ont recours les comités de lecture de revues scientifiques (Lefebvre). Dans le domaine de l'enseignement, l'EP est relativement courante au sein des cours universitaires à forts effectifs et des MOOC (Gamage, Staubitz et Whiting). L'EP s'est aujourd'hui simplifiée grâce au recours à des systèmes de gestion de l'apprentissage tels que Moodle LMS.
- 4 Les fondations théoriques de l'EP reposent sur les théories du développement cognitif de Piaget et la théorie socioculturelle du développement cognitif de Vygotsky. L'évaluation par les pairs appelle un travail collectif et semble de facto en adéquation avec les principes contemporains d'une pédagogie active revendiquée par les méthodologues des perspectives interactionniste, socio-culturelle ou socio-interactionniste.
- 5 Déjà en 1969, Burke s'intéressait à l'évaluation entre pairs et à l'auto-évaluation. Aujourd'hui, la littérature sur l'EP est d'une grande richesse. Les chercheurs ont proposé des définitions, des outils de conception, et souligné la diversité des modèles. La définition même des « pairs » a évolué (Falchikov). Elle désignait dans un premier temps les étudiants d'un âge et niveau d'étude proche (Goldschmid et Goldschmid). Whitman proposa ensuite des catégories plus subtiles : il nomma ainsi pairs proches (*near-peers*) les pairs plus expérimentés que les apprenants, et co-pairs (*co-peers*) ceux qui étaient au même niveau (Whitman). Topping révéla le caractère protéiforme de l'EP. À titre d'exemples, parmi les variables relevées au sein des EP figurent : le public concerné (enseignement primaire, secondaire, universitaire), les modalités (évaluation d'une ou de plusieurs productions ; évaluation en présentiel, à distance ; anonyme, non anonyme...) ou encore les objectifs (feedback qualitatif, notation chiffrée ; évaluation formative, sommative...) (Topping).
- 6 L'EP constitue ainsi un domaine vaste qui fait appel à d'autres champs d'études et concepts. Étudier l'EP implique une réflexion sur les outils numériques et la question des affordances (Blin) ; sur l'évaluation ; la rétroaction, objet d'étude en gestion des organisations depuis les années 1980 (Ramaprasad ; Hattie et Timperley) ; l'auto-régulation de l'apprentissage (Nicol et Macfarlane-Dick) ; ou encore le développement de la métacognition (Clayton Bernard et Kermarrec).

1.2. Bénéfices escomptés d'une évaluation par les pairs

- 7 L'EP soulève des interrogations légitimes portant principalement sur la compétence des pairs et la valeur des évaluations effectuées. Néanmoins, les atouts escomptés sont nombreux. La correction étant répartie entre plusieurs évaluateurs, elle est plus rapide qu'une évaluation traditionnelle et les délais d'attente du feedback sont par conséquent réduits. Cette efficacité rend l'EP suspecte aux yeux de ses critiques. La question sous-jacente de la compétence des pairs fait débat car le manque d'expertise des évaluateurs et une évaluation de mauvaise qualité peuvent avoir des conséquences négatives en limitant le souhait de l'étudiant de réviser son travail initial, en abaissant les standards ou encore en développant un sentiment de confiance excessif induit par des louanges non méritées (Cambre, Klemmer et Kulkarni). Cependant, Topping a montré la bonne qualité du feedback des pairs, en dépit du manque d'expertise de ces derniers en comparaison avec celle d'un enseignant expert. Ainsi, un pair moins compétent mais disposant de davantage de temps que l'enseignant peut effectuer une évaluation tout aussi fiable (Topping).
- 8 De nature quantitative ou qualitative, l'EP peut avoir une valeur formative importante en rendant possible la rétroaction malgré des effectifs importants. Nommée « feedback » en anglais, la rétroaction, afin d'être efficace (*effective feedback*), doit répondre à trois questions selon Hattie et Timberley : elle doit faire le lien avec les objectifs visés (*Where am I going ? - Feed Up*), permettre de regarder le chemin parcouru (*How am I going ? - Feed Back*) et, de façon plus prospective, le chemin qu'il reste à parcourir (*Where to next ? - Feed Forward*).
- 9 La rétroaction apportée rend l'évaluation utile pour l'étudiant-auteur. Elle permet ainsi de dépasser la simple *évaluation-sanction*. À la lueur de la définition proposée par Tabory, nous définissons l'évaluation-sanction comme visant à sanctionner un résultat, à confirmer, grâce à l'approbation d'une figure d'autorité, l'adéquation entre le résultat et les attendus. La validation est alors conditionnée par des normes, et la non-validation met principalement en exergue les difficultés rencontrées.
- 10 L'EP permet à l'étudiant-évaluateur d'adopter, pour sa part, une nouvelle perspective en tant que correcteur. Il est ainsi amené à justifier ses propos et trouver les mots justes, à porter un regard critique et faire preuve d'objectivité, ce qui renforce la compréhension du sujet et l'assimilation du cours. En faisant appel à des compétences diverses, l'EP peut occuper une place intéressante dans l'apprentissage, elle permet aussi aux étudiants de mieux se situer par rapport aux autres pour progresser. L'EP n'est pas nécessairement associée à l'auto-évaluation, néanmoins elle l'est dans de nombreux cas (Gamage, Staubitz et Whiting). Notre expérimentation fit usage de cette combinaison. L'auto-évaluation devait amener l'étudiant à mieux se situer par rapport aux objectifs visés (*Feed Up*), à porter un regard critique sur son travail (*Feed Back*), à se remettre en question et poursuivre une réflexion métacognitive. Ces objectifs nécessitent le recours à différentes compétences, ce qui nous porte à croire que l'EP et l'auto-évaluation seraient particulièrement adaptées à une évaluation par compétences.

2. Le contexte de notre évaluation par les pairs

2.1. Présentation de la formation

- 11 Cette évaluation s'adressait à un public d'étudiants inscrits en master 1 MEEF second degré, allemand, anglais, espagnol, ou créole à l'Université de la Réunion en 2021-2022, et suivant un élément constitutif d'unité d'enseignement (ECUE) intitulé « Enjeux didactiques de l'enseignement des langues ». Malgré de nombreux points communs, ces langues vivantes ont des traditions didactiques différentes. Divisé en deux parties, l'ECUE doit amener les étudiants non seulement à découvrir les notions clés de la didactique des langues et ses théories d'apprentissage, mais également à prendre connaissance du prescrit et de la didactique institutionnelle. Une réflexion sur la rupture et la continuité dans l'enseignement des langues a été menée : l'un des objectifs était d'amener les étudiants à mobiliser les connaissances acquises afin d'identifier certaines stratégies et techniques contemporaines d'enseignement des langues, de parvenir à retracer leurs origines et identifier les rémanences des méthodes, en particulier dans les techniques et outils. Les étudiants étaient amenés à voir que les méthodes obéissaient à des objectifs différents mais étaient rationnelles, que les paradigmes se répondaient et se complétaient. Cette réflexion sur la rupture et la continuité conduisit naturellement à une réflexion sur les modalités d'évaluation.

2.2. Contexte techno-pédagogique et modalités d'évaluation du cours

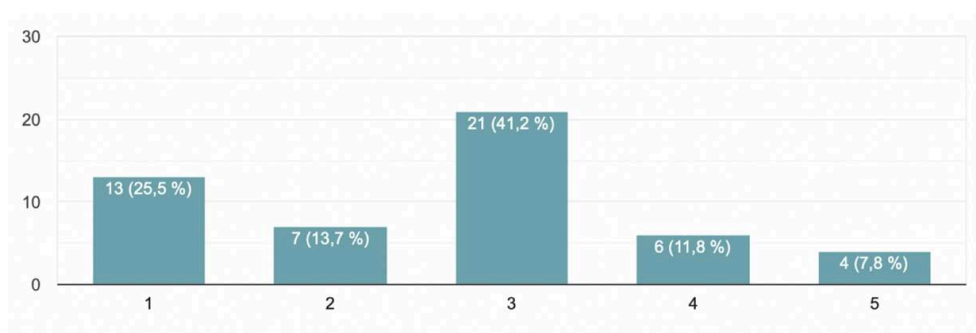
- 12 56 étudiants étaient inscrits et 54 ont effectué les évaluations assidûment (4 germanistes, 19 hispanistes, 28 anglicistes, 14 créolistes). Ce cours était initialement conçu comme un cours hybride alternant des cours en présentiel synchrones et des cours à distance asynchrones hébergés sur la plateforme numérique d'apprentissage Moodle. La capacité réduite de salles en demi-jauges nous contraignit à remplacer la plupart des séances synchrones en présentiel par des séances co-modales auxquelles certains étudiants assistaient en présentiel, et d'autres à distance via Zoom.
- 13 L'évaluation de l'ECUE contient six travaux. Une moyenne chiffrée devait être transmise dans le cadre de la validation du master 1. Le recours à six évaluations devait permettre de varier les modalités d'évaluation et les évaluateurs, d'évaluer des connaissances et des compétences, et ainsi essayer de tendre vers une évaluation la plus objective et complète possible. L'évaluation repose sur quatre questionnaires (coefficient 1) en ligne en temps limité visant une évaluation traditionnelle de connaissances (deux questionnaires formatifs, deux questionnaires sommatifs) ; une évaluation de « devoir maison » corrigée par l'enseignant (coefficient 2) ; une évaluation par les pairs associée à une auto-évaluation (coefficient 2). C'est cette dernière évaluation qui fait l'objet de la présente analyse.

3. Les phases de notre évaluation par les pairs

- 14 Nous avons conduit une enquête préliminaire dès le début du cours afin d'avoir une meilleure connaissance des outils et des conditions de travail des étudiants, de leurs attentes et de leur expérience. 51 réponses ont été obtenues et ont montré que seuls

58,8 % des étudiants avaient déjà suivi des formations hybrides les années précédentes. À la question « Êtes-vous a priori favorable aux travaux à effectuer en ligne et corrigés par des pairs ? », 25,5 % des étudiants se disaient très favorables (figure 1). 29 étudiants (63 %) ont répondu positivement à la question « Si les travaux à effectuer en ligne et corrigés par des pairs étaient notés, souhaiteriez-vous que cette note soit prise en compte dans l'évaluation de l'UE ? ». 17 étudiants y étaient défavorables (37 %).

Figure 1. Enquête préliminaire : « Êtes-vous a priori favorable aux travaux à effectuer en ligne et corrigés par des pairs ? » (1- Oui tout à fait ; 5- Non pas du tout)



- 15 Destinée à de futurs enseignants, l'évaluation consistait à produire une fiche d'activité conçue pour des élèves du secondaire, comportant des exercices d'entraînement en réception écrite ou orale répondant aux critères vus en cours, et en joignant de brèves explications. Le cours étant inter-langues, les pairs-évaluateurs pouvaient ne pas être spécialistes de la même langue que l'étudiant-auteur. C'est pourquoi une traduction en français des consignes aux élèves était requise.
- 16 L'évaluation a été menée à l'aide du module « Atelier » disponible sur la plateforme Moodle utilisée par l'établissement. Le module n'a pas été modifié, et les quatre phases prédéfinies dans le module d'activité ont été suivies :
1. La phase de mise en place et de configuration, lors de laquelle ont été saisies les consignes (annexe 1 – Fiche consignes). Le formulaire d'évaluation a été préparé pendant cette phase puis transmis aux étudiants. Le formulaire d'évaluation comportait 7 aspects (annexe 2 – Critères d'évaluation). Nous avons paramétré le formulaire de façon à garantir l'anonymat des personnes évaluées : les évaluateurs ne devaient par conséquent pas avoir connaissance de l'identité des pairs qu'ils évaluaient. La stratégie d'évaluation retenue était « l'évaluation cumulative », permettant de joindre commentaires et notes.
 2. La phase de remise, pendant laquelle les étudiants ont rédigé puis remis leurs travaux en ligne. Les étudiants ont disposé d'un mois (19/11-18/10) et la durée estimée de cette phase était de 2 h 30-3 h. À l'issue de cette phase, nous avons attribué trois travaux à évaluer aux étudiants.
 3. La phase d'évaluation a duré 10 jours (18/10-28/10). La durée estimée de cette phase était d'1 h-1 h 30. Les étudiants ont procédé à l'évaluation de trois pairs et à leur auto-évaluation ; un forum de questions a été ouvert pendant cette période. Ce forum permettait aux étudiants de poser toute question concernant le processus d'évaluation, les critères d'évaluation ou encore les difficultés rencontrées ; ce forum était modéré par l'enseignante.
 4. La phase de notation des évaluations a permis le contrôle des évaluations par l'enseignante, la réévaluation de certaines copies puis le calcul de la note finale. La note finale était composée de la moyenne des évaluations des pairs (coefficient 3) et de l'auto-évaluation (coefficient 3). Le module « Atelier » attribue automatiquement deux notes aux étudiants : une note pour leur travail (calculée à partir des évaluations faites par les pairs et par eux-

même), ainsi qu'une note pour la qualité des évaluations qu'ils ont fournies. Cette seconde note est basée sur la comparaison des évaluations d'un même devoir par plusieurs étudiants. Le formateur peut aussi attribuer manuellement cette note. Il s'agit donc d'un élément permettant au formateur de visualiser les écarts et d'avoir une indication quant à la qualité des évaluations faites par les pairs. Nous avons fait le choix de ne pas inclure cette note attribuée automatiquement par Moodle dans le calcul de la moyenne. En revanche, nous avons conçu cette note comme un indicateur afin de vérifier voire procéder à la nouvelle notation des copies n'ayant pas obtenu de note de qualité d'évaluation satisfaisante.

4. Bilan

4.1. Résultats des auto-évaluations et évaluations par les pairs

- 17 55 étudiants ont effectué l'activité intégralement et ont reçu une note. Seul un étudiant a remis son devoir mais n'a pas procédé à temps à l'évaluation de ses pairs. La moyenne des évaluations des pairs est de 15,7/20. La moyenne des auto-évaluations est de 16,7/20. L'écart moyen entre les auto-évaluations et l'évaluation des pairs est donc d'un point, les auto-évaluations étant légèrement plus hautes. 2 étudiants ont obtenu exactement la même note à l'évaluation par les pairs et à l'auto-évaluation. 21 étudiants se sont attribué une note inférieure à celle qui était donnée par les pairs. Pour 8 d'entre eux, l'écart entre les deux notes est faible (entre 0,1 et 1,5 points). Pour 13 d'entre eux, l'écart est de 1,6 à 2,5 points. On relève que les quatre étudiants ayant obtenu une note de 19 ou 20 des pairs s'accordent à eux-mêmes une note inférieure. 32 étudiants se sont attribué une note supérieure à celle qui était délivrée par les pairs. Pour 16 étudiants, l'écart de notes était faible (entre 0,1 et 2 points) ; 12 étudiants s'attribuaient une note supérieure de 2,1 à 5 points à celle des pairs. Enfin 4 étudiants se sont octroyé une note supérieure de plus de 5 points à celle des pairs. Lorsque l'écart de notes entre l'auto-évaluation et la note des pairs, ou entre les trois notes des pairs, était trop important, les copies ont été renotées par l'enseignante. 6 copies ont été réévaluées. Alors qu'en 1969 Burke relevait que les étudiants ne parvenaient pas à s'évaluer correctement et avaient tendance à s'attribuer de meilleures notes que méritées, nous constatons une tendance similaire lors de notre évaluation pour 32 étudiants. Néanmoins on relève toutefois que la propension à se surévaluer paraît être inverse pour les meilleurs étudiants.

4.2. Retour d'expérience de l'enseignante et des étudiants

- 18 Au regard des notes et des commentaires effectués par les pairs lors des évaluations, le bilan que nous dressons est, dans l'ensemble, positif. L'évaluation par les pairs a rendu possible un travail inter-langues entre différents groupes ; l'engagement des étudiants dans la rédaction des commentaires s'est révélé satisfaisant. Il serait nécessaire de mener des recherches scientifiques afin de s'en assurer, mais les évaluations témoignent à première vue d'efforts importants des étudiants afin de s'appuyer sur le cours et de proposer une rétroaction argumentée. Les notes des pairs paraissent cohérentes, en revanche l'écart entre les notes des pairs et l'auto-évaluation semble révéler que l'auto-évaluation est plus délicate ; elle n'en demeure pas moins utile car elle permet non seulement aux étudiants de prendre conscience des écarts, mais elle permet une rétroaction plus complète, permettant les *Feed Back*, *Feed Up* et *Feed Forward*

définis par Hattie et Timberley. En ce sens, l'auto-évaluation paraît faciliter une analyse métacognitive. Aussi ce n'est pas l'EP seule qui confère son intérêt à l'activité, mais plutôt la complémentarité entre l'évaluation par les pairs et l'auto-évaluation.

- 19 Néanmoins le bilan n'est pas entièrement positif. Nous avons rencontré des difficultés techniques dans le calcul des notes, et comme le remarquait Falchikov, on relève aussi une quantité de travail notable, à la fois en amont (préparation de la grille d'évaluation, paramétrages), pendant l'activité (supervision, accompagnement, rappel des échéances) et en aval (évaluation des notations, réévaluations, ajout du feedback du formateur).
- 20 Une enquête anonyme effectuée à l'issue de l'EP a permis d'obtenir le retour des étudiants sur cette expérience pédagogique. Les 23 résultats recueillis montrent qu'à l'issue de cette première expérience, la plupart des étudiants sont favorables à une EP notée (43,5 % y sont très favorables, 26,1 % favorables, 8,7 % sont neutres, 13 % défavorables et 8,7 % très défavorables). Une question ouverte permettait aux étudiants de justifier leurs choix ; il semblerait que la question de l'anonymat offre une explication partielle à ces chiffres. En effet, plusieurs étudiants avaient inscrit leur nom dans le corps de la copie.
- 21 On relève aussi que tous les étudiants concernés déclarent avoir consulté le barème avant de rédiger et déposer le devoir. La fiabilité des notes semble admise puisque tous les étudiants déclarent que les notes paraissent justes : 52,2 % pensent que la moyenne des notes données par les pairs semble « tout à fait juste » et 47,8 % « dans l'ensemble, juste ». En cas de décalage important entre la note des pairs et l'auto-évaluation, tous les étudiants disent être favorables à une nouvelle notation par l'enseignant. L'égalité de poids accordée à la note des pairs et à la note de l'auto-évaluation est satisfaisante pour 91,3 % étudiants. Néanmoins, 2 étudiants (8,7 %) auraient souhaité que l'auto-évaluation ait un plus fort coefficient. Enfin, la durée des phases de l'activité semble avoir été sous-estimée. En effet, 34,8 % des étudiants déclarent avoir passé le même temps que la durée indicative (2 h 30/3 h) afin de concevoir et déposer leur devoir ; 65,2 % des étudiants déclarent avoir passé plus de temps que la durée indicative, et aucun n'a passé moins de temps. Pour la phase d'évaluation, 56,5 % des étudiants indiquent avoir passé le même temps à évaluer les activités des pairs que la durée indicative (1 h, 1 h 30), 39,1 % déclarent avoir passé plus de temps, et un seul étudiant (4,3 %) déclare avoir passé moins de temps.

5. Points de vigilance et pistes d'améliorations

5.1. Travail collectif et accompagnement des enseignants

- 22 À l'issue de cette première évaluation par les pairs, nous émettons quelques recommandations. Dans un premier temps, un travail collectif impliquant enseignants, ingénieurs pédagogiques et informaticiens devrait permettre de faciliter les étapes de conception, de mise en œuvre ainsi que le suivi technique de l'activité sur la plateforme.
- 23 Au-delà des questions d'ingénierie de la formation et de pédagogie, les aspects logistiques et techniques sont complexes et requièrent un accompagnement fort de la part de l'institution. Les modules de base de l'activité Moodle sont en effet

contraignants et paraissent pour l'instant faiblement utilisés au sein de l'établissement, ce qui explique que leurs spécificités soient encore relativement méconnues.

5.2. « L'évaluation cumulative »

- 24 Si des modules additionnels semblent exister, un choix restreint de modalités d'évaluation est proposé dans le logiciel Moodle basique. Après avoir mené quelques recherches sur la question et consulté des ingénieurs pédagogiques et de la documentation technique, nous avons opté pour l'« évaluation cumulative ». Il s'agit d'un type d'évaluation qui permet de recourir à des commentaires et de donner une note sur différents aspects (annexe 2 – Critères d'évaluation). Nous avons ainsi proposé une grille d'évaluation composée de 7 items et permettant d'obtenir de 2 à 6 points selon les aspects. Chaque item avait une pondération de 1 et la somme des points était de 20. Cependant, une fois les évaluations effectuées par les pairs, les notes ont été calculées automatiquement selon un autre mode de calcul dépendant de la plateforme. Les notes des différents items ont été normalisées dans une plage allant de 0 % à 100 %. La note totale a ensuite été calculée en effectuant la moyenne des notes normalisées. Afin d'obtenir un calcul de note totale correspondant à la somme des points obtenus, il aurait donc fallu modifier les pondérations de chaque aspect. Néanmoins, à ce stade, il n'était techniquement plus possible de modifier les critères initiaux sans modifier le travail effectué par les étudiants. Par chance, ce système de calcul s'est révélé à l'avantage de la majorité des étudiants ; seuls quelques étudiants (ceux qui avaient les meilleurs résultats) ont été légèrement désavantagés. Nous avons précédemment relevé que les meilleurs étudiants avaient obtenu une note d'auto-évaluation inférieure à la note des pairs ; on pourra se demander si ces étudiants ont tendance à se sous-évaluer ou ont pâti de ce mode de calcul de la note.

5.3. Travail collectif et anonymat des copies

- 25 Relativement novices dans l'EP voire dans l'évaluation, plusieurs étudiants ont rencontré des difficultés pour évaluer leurs pairs. Aussi les questions de la responsabilité et de la légitimité ne peuvent pas être négligées. Ce travail entre pairs doit donc s'accompagner d'une réflexion sur les compétences, la bienveillance et la compétition. S'il s'agit *in fine* d'exploiter le travail entre pairs afin de faciliter les apprentissages, une réflexion sur la coopération (ici entre étudiants de M1 allemand, anglais, créole, espagnol, spécialistes de langues vivantes différentes), la collaboration (entre experts spécialistes d'une même langue) voire la « coopération » (coopération avec des concurrents) pourra être envisagée.
- 26 Il s'agira de surcroît de veiller à l'anonymat des copies. Nous avons effectué un paramétrage spécifique afin que les copies soient anonymes, malheureusement plusieurs étudiants ont inscrit leur nom et numéro d'étudiant dans le corps du fichier soumis. Si l'EP donne lieu à une note, l'anonymat nous semble requis afin de garantir une certaine objectivité et libérer les étudiants-évaluateurs de tout sentiment de culpabilité.

5.4. L'accompagnement des étudiants-évaluateurs

- 27 À la lueur des modèles de conception d'EP proposés par les chercheurs (Falchikov ; Gielen et De Wever), il nous semblerait nécessaire de guider davantage les étudiants dans leur travail d'évaluation des pairs. Pour ce faire, nous proposons d'inclure une phase d'analyse de travaux exemplaires, sinon de co-construction de la grille d'évaluation (Clayton Bernard et Kermarrec ; Gielen et De Wever), ce qui pourrait permettre en sus de conférer aux étudiants-correcteurs davantage de légitimité. L'analyse d'exemple ne devrait pas présenter de difficultés techniques car il s'agit d'ores et déjà d'une option disponible dans le module d'activité Moodle. Cette analyse gagnera à être complétée par un travail sur les modalités d'administration de feedback, permettant de réfléchir à la façon dont il est possible pour l'étudiant-correcteur d'apporter des précisions, des pistes d'amélioration, d'encourager un travail réflexif et de permettre *in fine* à l'étudiant-auteur d'améliorer le travail initialement proposé.

Conclusion

- 28 Cette expérimentation a soulevé plusieurs interrogations et nous amène à identifier des pistes de travail : il s'agit des liens entre évaluation par les pairs et apprentissage ; entre travail collectif et apprentissage ; des écarts de notes entre auto-évaluation et évaluation par les pairs ; et de l'étayage apporté aux évaluateurs. Nous rappelons que ces champs d'études sont d'ores et déjà investis par les chercheurs.
- 29 Le recours à plusieurs dispositifs d'évaluation au sein d'un même cours a permis l'intégration de l'évaluation de différents éléments (connaissances, compétences) et avec des objectifs variés (formatif, sommatif, visant la rétroaction). La combinaison d'une évaluation par les pairs et d'une auto-évaluation s'est révélée positive et semblerait conduire à un enrichissement, bien que ce travail exploratoire n'apporte pas suffisamment d'éléments afin de le confirmer. Si l'utilité d'une évaluation par les pairs a été démontrée par les chercheurs, cela ne signifie pas pour autant qu'une réflexion sur l'ingénierie pédagogique ne soit plus requise.

BIBLIOGRAPHIE

Blin, Françoise. « The Theory of Affordances. » *Language-Learner Computer Interactions: Theory, methodology and CALL applications*, sous la direction de Catherine Caws et Marie-Josée Hamel, John Benjamins Publishing Company, 2016. DOI : 10.1075/lss.2.

Burke, Ronald J. « Some Preliminary Data on the Use of Self-Evaluations and Peer Ratings in Assigning University Course Grades. » *The Journal of Educational Research*, vol. 62, n° 10, Taylor et Francis, Ltd., 1969, pp. 444-448. <http://www.jstor.org/stable/27532261>.

Cambre, Julie, Scott Klemmer, et Chimnay Kulkarni. « Juxtapeer: Comparative Peer Review Yields Higher Quality Feedback and Promotes Deeper Reflection. » *Proceedings of the 2018 Conference on*

Human Factors in Computing Systems (CHI '18), Association for Computing Machinery, New York, 2018. DOI : 10.1145/3173574.3173868.

Clayton Bernard, Rebecca, et Gilles Kermarrec. « L'évaluation par les pairs et le feedback vidéo : un dispositif pour développer l'autorégulation et la régulation par partage social en cours de langue. » *Les Annales de QPES*, vol. 1, n° 1, juin 2020, pp. 1-21. DOI : 10.14428/qpes.v1i1.55783.

Condette, Jean-François. « Entre mesure de l'activité scientifique par les pairs et contrôle institutionnel. L'évaluation des enseignants-chercheurs des facultés des lettres sous la III^e République (1870-1940). » *Spirale, Revue de recherches en éducation, L'évaluation des enseignants histoire, modalités et actualités*, n° 49, 2012, pp. 83-105. DOI : 10.3406/spira.2012.1110.

Falchikov, Nancy. *Learning Together, Peer Tutoring in Higher Education*. Routledge Falmer, 2001.

Gamage, Dilrukshi, Thomas Staubitz, et Mark Whiting. « Peer assessment in MOOCs: Systematic literature review. » *Distance Education*, vol. 42(2), 2021, pp. 268-289. DOI : 10.1080/01587919.2021.1911626.

Gielen, Mario, et Bram De Wever. « Structuring the Peer Assessment Process: a Multilevel Approach for the Impact on Product Improvement and Peer Feedback Quality. » *Journal of Computer Assisted Learning*, 2015-10, vol. 31(5), pp. 435-449. DOI : 10.1111/jcal.12096.

Glaudes, Pierre. « L'évaluation de la production scientifique en France par l'Agence d'évaluation de la recherche et de l'enseignement supérieur. » *Mélanges de la Casa de Velázquez*, n° 44-2, 2014, pp. 293-300. DOI : 10.4000/mcv.5858.

Goldschmid, Barbara, et Marcel L. Goldschmid. « Peer Teaching in Higher Education: A Review. » *Higher Education*, (5), 1976, pp. 9-33. DOI : 10.1007/BF01677204.

Hattie, John, et Helen Timperley. « The Power of Feedback. » *Review of Educational Research*, vol. 77, n° 1, 2007, pp. 81-112. DOI : 10.3102/003465430298487.

Lefebvre, Muriel. « L'évaluation des savoirs scientifiques : modalités et enjeux. » *La publication scientifique : analyses et perspectives*, sous la direction de Joachim Schöpfel, Hermès, 2008, pp. 299-316. https://archivesic.ccsd.cnrs.fr/sic_00429967/document.

Nicol, David J., et Debra Macfarlane-Dick. « Formative Assessment and Self-regulated Learning: A Model and Seven Principles of Good Feedback Practice. » *Studies in Higher Education*, 31(2), 2006, pp. 199-218. DOI : 10.1080/03075070600572090.

Picard, Emmanuelle. « Les enseignants-chercheurs : une évaluation centralisée (1873-1992). Du comité consultatif de l'enseignement supérieur (CNU) (1873-1992). » *Spirale. Revue de recherches en éducation, L'évaluation des enseignants histoire, modalités et actualités*, n° 49, 2012, pp. 69-82. DOI : 10.3406/spira.2012.1109.

Ramaprasad, Arkalud. « On the Definition of Feedback. » *Behavioral Science*, vol. 28, 1983, pp. 4-13. DOI : 10.1002/bs.3830280103.

Tabory, Marc. « Notation-Sanction : le bât blesse... » 2006. <https://www.meirieu.com/ECHANGES/taboryevalu.pdf>.

Topping, Keith J. « Peer Assessment: Learning by Judging and Discussing the Work of Other Learners. » *Interdisciplinary Education and Psychology*, 1(1):7, 2017, pp. 1-17.

Whitman, Neal. *Peer Teaching. To Teach Is to Learn Twice*. ASHE-ERIC Higher Education Report No. 4. Washington, DC, Association for the Study of Higher Education, 1988. <https://files.eric.ed.gov/fulltext/ED305016.pdf>.

Annexe 1. Fiche consignes

Cette activité est liée à ce que vous allez faire et voir en stage. Cette activité est inter-langues, vous pourrez ainsi être amenés à évaluer des travaux d'étudiants spécialistes d'une autre langue.

1° Dans un premier temps il s'agit de déposer une « fiche de dépôt » pour le 18 octobre ; elle comportera une activité d'entraînement et de brèves explications de votre part.

- Pensez qu'il s'agit d'une activité inter-langues, et que vos évaluateurs ne sont pas spécialistes de votre LV. Les consignes seront traduites en français.

- L'activité est une activité d'entraînement en réception orale ou écrite, destinée à la classe de votre choix, dans la langue de votre choix. Cette activité doit être une activité d'entraînement répondant aux critères vus en cours.

- Elle comportera environ 5 exercices/questions.

- Vos explications justifieront le fait qu'il s'agit d'une activité conçue dans le cadre d'un entraînement et non d'une évaluation.

2° Dans un second temps vous évalueriez les activités proposées par trois pairs avant le 28 octobre.

- Il faudra s'assurer qu'il s'agit bien d'une activité conçue dans le cadre d'un entraînement, et que l'activité n'est donc pas évaluative.

- Vous remplirez la grille d'évaluation pour chaque activité que vous évaluez.

3° Vous procéderez ensuite à votre auto-évaluation avant le 28 octobre.

Annexe 2. Critères d'évaluation

Les items suivants permettent de justifier que l'activité soumise a bien été conçue comme un entraînement pour les élèves et qu'elle n'est donc pas à visée évaluative.

Aspect formel :

- Le document support est bien inclus (il est déposé ou un lien fonctionnel est donné) (... /1)

- La fiche d'entraînement est bien déposée (.../1)

- La fiche est correctement remplie, l'aspect formel est recevable (pas d'erreurs de langue) (.../1)

Ancrage culturel :

- Le support choisi répond à la commande institutionnelle (le niveau choisi est mentionné, le support respecte les programmes du niveau choisi) (.../1)

- Les justifications apportées sont claires et cohérentes (.../1)

Choix didactique : Une réflexion a été menée quant aux éléments qui font l'objet d'une attention particulière. Il ne s'agit pas de viser une compréhension exhaustive ; des points sont ciblés, il n'est pas question de décortiquer.

- Il est demandé à l'élève d'identifier/repérer un nombre raisonnable d'informations. (.../1)

- Les justifications apportées sont claires et cohérentes (.../1)

Modalités :

- Les modalités retenues sont adaptées aux consignes (.../1)

- Le nombre d'écoutes et de pauses est raisonnable (..../1)
- Les justifications apportées sont claires et cohérentes (..../1)

Cohérence interne : Progression du plus simple au plus complexe, de l'explicite vers l'implicite, du connu vers l'inconnu...

- Les activités doivent être effectuées dans l'ordre ; elles ne sont pas interchangeables (..../1)
- Les justifications apportées sont claires et cohérentes (..../1)

Les questions :

La plupart des questions posées constituent une aide : il ne s'agit pas de vérifier que l'élève a compris mais de l'aider à mieux comprendre, de le guider vers l'accès au sens. (..../2)

Stratégies : À l'issue de l'activité, l'élève dispose de deux nouveaux outils méthodologiques / stratégies (voire plus).

- Présence de deux stratégies (..../2)
- Les deux stratégies sont bien adaptées aux spécificités du document (..../2)
- Les stratégies font l'objet d'une réflexion et d'explications aux élèves (en aval, en amont, ou pendant l'activité). (..../2)

NOTES

1. L'une des études menées par Burke en 1969 à l'Université du Minnesota visait à accroître la participation des étudiants dans l'attribution des notes. À l'issue des auto-évaluations et évaluations par les pairs qu'il mena, Burke conclut que les étudiants ne parvenaient pas à s'attribuer une note de manière réaliste. En revanche l'évaluation par les pairs était cohérente, donnait lieu à une distribution des notes plus typique, et concordait avec les évaluations des enseignants.

RÉSUMÉS

L'objet de cette note de pédagogie universitaire est une évaluation expérimentale par les pairs. Le contexte est une formation en didactique des langues à l'Institut national supérieur du professorat et de l'éducation (Inspé) de La Réunion en 2021-2022, s'adressant à cinquante-six étudiants. L'étude s'appuie sur deux enquêtes menées auprès des participants en amont puis en aval de l'évaluation par les pairs, qui constitue ainsi un cas concret. Après avoir présenté la pertinence et les atouts escomptés du dispositif en nous appuyant sur les travaux scientifiques qui traitent de ce point, nous concevons cette activité comme étant bien davantage qu'une simple « évaluation-sanction ». Dans notre étude de cas, l'évaluation par les pairs est associée à une auto-évaluation, elle occupe une place certaine dans l'apprentissage, elle facilite la rétroaction et permet notamment une réflexion métacognitive. À l'issue de l'expérimentation, plusieurs interrogations sont soulevées, comme l'écart entre les notes des auto-évaluations et des

évaluations par les pairs. Nous terminons notre étude par des points de vigilance et des recommandations.

This note on language teaching in higher education deals with an experimental peer assessment activity. The context is set within a course in Language didactics at the School of Education of Reunion Island (Inspé) attended by around fifty students. The study relies on two enquiries carried out before and after the concrete case. Having mentioned the relevance and the potential benefits of peer-assessment while summoning research, we consider this activity as much more than a merely summative assessment. Associated here with self-assessment, peer assessment incontestably plays a part in the learning process, it facilitates feedback, and allows for metacognitive analysis. Following the experiment, many questions are asked, such as grade differences between self-assessment and peer-assessment. We eventually close this study with cautionary notes and recommendations.

INDEX

Keywords : collaborative learning, feedback, higher education pedagogy, language didactics, learning, peer assessment, testing (teaching methods)

Mots-clés : apprentissage, didactique des langues, évaluation (méthodes pédagogiques), évaluation par les pairs, pédagogie universitaire, rétroaction, travail collaboratif

AUTEUR

GUILÈNE RÉVAUGER

Guilène Révauger est maîtresse de conférences en études du monde anglophone à l'Institut national supérieur du professorat et de l'éducation (Inspé) de l'Université de La Réunion. Auparavant professeure agrégée d'anglais puis formatrice académique, elle a été amenée à enseigner en Guyane française et en Polynésie française, dans le second degré ainsi qu'en classe préparatoire aux grandes écoles et à l'université.

Elle est l'auteur d'une thèse en civilisation britannique intitulée *Fonctions, pouvoirs et influences d'un acteur de la politique étrangère britannique : le Foreign and Commonwealth Office (1968-1985)*.

Aujourd'hui membre du laboratoire DIRE (Déplacements, Identités, Regards, Écritures), ses recherches portent actuellement sur la didactique de l'anglais, le plurilinguisme, l'utilisation du numérique dans l'enseignement/apprentissage des langues et la formation des enseignants.

guilene.revauger[at]univ-reunion.fr