# Quand le numérique facilite la correction des questions d'examens : expérimentation de Gradescope à l'UCLouvain

Yves Deville, Christine Jacqmot, Olivier Pereira
UCLouvain, Belgique
{Yves.Deville, Christine.Jacqmot,Olivier.Pereira}@uclouvain.be

## Résumé :

La correction d'examens est coûteuse en temps, en particulier pour les cohortes à effectifs élevés. Des solutions numériques existent pour faciliter, voire automatiser la correction des travaux des étudiants et l'UCLouvain en utilise quelques-uns (dont les deux systèmes OpenSource Moodle ou Inginious, par exemple). Ces systèmes n'apportent cependant pas de solution pour les nombreux enseignants de notre institution organisant des examens écrits, avec questions ouvertes, organisés en amphithéâtre, où le papier reste de mise. Temps de correction, fiabilité et stabilité de l'évaluation, objectivation de la note, rétroaction … tels sont des exemples de défis que rencontrent tous les enseignants confrontés à la correction de telles épreuves, en particulier pour de grandes cohortes.

Dans le cadre de l'Université numérique, programme lancé il y a 5 ans, l'UCLouvain a investi dans un logiciel procurant un soutien aux enseignants dans la correction des examens écrits à réponses ouvertes ou schémas sur papier : Gradescope. Développé à l'Université de Californie à Berkeley, utilisé par des enseignants de 500 universités, l'UCLouvain est la première université hors Etats-Unis à avoir parié sur les plus-values de cette solution pour en faire un outil institutionnel. Le processus est à la fois simple et puissant : scan des copies, critères objectivés de correction par question, précorrection de certains types de questions par un module reposant sur l'IA, attribution des points et calcul de la note finale selon les critères identifiés par le ou les correcteurs, accès aux copies corrigées par les étudiants, etc. (Singh, 2017). Dans cette communication nous présenterons (a) le contexte de ce projet pilote, la problématique qu'il adresse et les atouts escomptés, tant pour les enseignants que pour l'institution, au regard de la qualité des évaluations, (b) les bilans tirés des premières exploitations par les enseignants, l'impact sur leurs pratiques ainsi que les perspectives.

Mots-clés : évaluation des acquis d'apprentissage, outil numérique, soutien aux enseignants, qualité, impact

## Contexte : Quand le numérique vient en renfort de la qualité

L'évaluation des acquis des étudiants est un moment fort de l'année académique, notamment par sa complexité (Berthiaume, 2011). Bien que chacun et chacune soit conscient de l'importance de varier les formats (Jacobs, 2004), on constate un recours fréquent aux QCM, pour leur facilité et rapidité de correction automatisée, alors que les enseignants se plaignent de leur inadéquation. L'UCLouvain s'est tournée vers le logiciel Gradescope, misant sur un certain nombre d'atouts de l'outils, en lien avec la qualité des évaluations :

### *Gain de temps*

Gradescope argumente d'un gain de temps important de lecture des réponses et de leur correction, grâce à la numérisation du processus (Singh, 2017). Mais on peut aussi souligner la facilité d'une correction menée de front par plusieurs correcteurs simultanément, traitant chacun un lot de questions, à partir de leurs ordinateurs respectifs : les copies, les critères et le système d'attribution de points sont instantanément partagés entre les correcteurs. Quel en est l'impact ?

### *Objectivité*

Le recours à une grille d'évaluation permet de définir les qualités attendues des productions des étudiants ou encore les réponses acceptées. Expliciter des critères et des indicateurs, préciser les indices observables et exprimer des niveaux de performance réduisent la subjectivité de la correction et de la note et évitent au correcteur de se laisser « distraire » par des éléments non essentiels (Mottiez-Lopez, 2016) (Berthiaume, 2011). La définition de critères, associés à des points « gagnés » ou « perdus », est au cœur du processus de correction dans Gradescope, dans une approche cependant plus analytique (Dawson, 2017) que qualitative (Biggs, 1999), (Berthiaume, 2011). Comment ce soutien à l'objectivité est-il perçu ?

### *Stabilité et homogénéité*

Un risque connu, lorsque plusieurs personnes participent à la correction ou lorsque les corrections prennent plusieurs jours, est de constater une variabilité de la sévérité de correction et de la note accordée entre correcteurs ou en fonction du moment et contexte de la correction. (Jacobs, 2004) (Hativa, 2000). Corriger question par question contribue, par exemple, à une application plus homogène de critères de correction.

### *Ajustement du barème*

Les critères de correction peuvent évoluer en cours de correction, en réponse à des contenus inattendus et au regard des contributions individuelles des étudiants (Mottiez-Lopez, 2016). Il est important que l'enseignant puisse ajuster son référentiel de correction en fonction de la situation qu'il rencontre. Mais avec des copies papier, la tâche est fastidieuse, et donc parfois évitée.

### *Rétroaction pour les étudiants*

Au-delà de la notation, on connaît l'intérêt des rétroactions qualitatives (Leduc, 2016). L'organisation de la consultation des copies papier annotées peut être très lourde, surtout

pour de grands groupes d'étudiants. Avec Gradescope, ces derniers peuvent consulter en ligne les critères de correction, leur copie corrigée, leur note ainsi que les commentaires rédigés par l'enseignant. Ils peuvent à leur tour adresser une notification aux enseignants, s'ils ne comprennent pas un élément, estiment avoir été mal compris, souhaitent prendre rendez-vous, etc.

### *Rétroaction pour l'enseignant*

La qualité et les détails des statistiques fournies par le logiciel à l'enseignant (pour la cohorte entière, par question, par critère… ) sont autant de sources d'informations sur la maîtrise de concepts par les étudiants, qui peuvent être exploitées lors de l'enseignement.

## Bilan et perspectives

Dans la section précédente, nous avons identifié six facteurs en lien avec la qualité des évaluations des acquis des étudiants. Qu'en est-il dans la pratique ? Notre expérience porte sur une cinquantaine de cours, trois sessions d'examens, et un total de plus de 10.000 copies d'examens. Une enquête menée auprès des enseignants de l'UCLouvain laisse présager que le recours à Gradescope est un réel soutien à la qualité des évaluations :

Le gain de temps est unanimement souligné, y compris pour les petites cohortes : entre 25% et 50% pour la moitié des enseignants et plus de 50% pour un quart des sondés. Ceci est attribué aux raisons suivantes :

- La numérisation du processus ;
- L'usage de critères de corrections objectivés, en permanence présents sur l'écran au moment de la correction ;
- Le caractère ajustable à tout moment des critères et des notations associées et la simplicité du système de critères (critères multiples avec points qui s'additionnent dans la majorité des cas) : Gradescope propage alors automatiquement l'adaptation à toutes les copies déjà corrigées, garantissant une équité de traitement sans consommation de temps ;
- La navigation par question est ensuite identifiée comme facteur d'efficience ;
- Vient ensuite la correction simultanée par plusieurs correcteurs.

Les conséquences sont intéressantes : retour aux QROC, schémas ou questions ouvertes et augmentation de la fréquence de tests formatifs.

Les trois facteurs principaux vecteurs de qualité mis en évidence dans l'enquête sont :

- L'explicitation des critères de correction, qui objective les attendus et qui permet un travail pédagogique avec les étudiants ;
- La stabilité de la correction au fil du temps grâce (1) aux critères de correction et les notations associées affichés en regard de la réponse corrigée et (2) à la correction question par question facilitée par l'interface ; la combinaison de ces deux éléments confère aux corrections stabilité et homogénéité (Hativa, 2000). Les enseignants parlent, eux, d'équité : la correction d'une question n'est pas influencée par la réponse à la question précédente, par exemple ;
- La rétroaction pour les étudiants est appréciée par les enseignants et par les étudiants tant pour la transparence (les enseignants parlent de « correction documentée » et de « notation étayée ») que pour la possibilité offerte aux étudiants d' «apprendre de

leurs erreurs ». Néanmoins, tous les enseignants n'osent pas encore utiliser cette fonctionnalité lors d'évaluations certificatives, par crainte d'être confrontés à une avalanche de récriminations infondées.

Plus de 90% des enseignants sondés souhaitent continuer à utiliser Gradescope et songent à en recommander l'usage à leurs collègues, même si quelques fonctionnalités manquent à l'outil (anonymat des copies, possibilité de double-correction, …).
L'enjeu sera de dépasser le stade du projet pilote, d'étendre l'usage de l'outil lorsqu'il apporte une plus-value et d'accompagner les enseignants qui n'oseraient pas faire le pas du numérique.

## Bibliographie


Biggs, J. (1999). Teaching for quality learning at university, Society for Research into Higher Education, 1999.

Berthiaume, D., David, J., David, T. (2011). Réduire la subjectivité lors de l'évaluation des apprentissages à l'aide d'une grille critériée : repères théoriques et applications à un enseignement interdisciplinaire, *Revue internationale de pédagogie de l'enseignement supérieur* 27(2), 2011, (consulté en septembre 2019). http://journals.openedition.org/ripes/524.

Dawson, P. (2017). Assessment rubrics: towards clearer and more replicable design, research and practice, *Assessment and evaluation in higher education*, 42(3), p347-360, DOI: 10.1080/02602938.2015.1111294.

Hativa, N. (2000). Assessing Student Learning. Dans *Teaching for effective Learning in Higher Education*, Springer, 2000, p307-325.

Jacobs, L.C. (2004) How to write better tests, Indiana University (consulté en septembre 2019). http://hdl.handle.net/10333/9104

Leduc, L. e.a. (2016). Viser le développement des pratiques d'Assessment for Learning (AfL) dans le contexte dela1ère année à l'université : cadre théorique, dispositif et initiatives pédagogiques résultant du projet « Feedbacks 1er bac ». Dans *L'évaluation à la lumière des contextes et des disciplines*, De Boeck Supérieur, 2016, p221-255.

Mottiez-Lopez, L. (2016). Une modélisation pour appréhender la référentiation dans l'évaluation des apprentissages des élèves. Dans *L'évaluation à la lumière des contextes et des disciplines*, De Boeck Supérieur, 2016, p169-192

Singh, A. e.a. (2017). Gradescope : a Fast, Flexible, and Fair System for Scalable Assessment of Handwritten Work. P*roceedings of the Fourth ACM Conference on Learning @ Scale,* 2017, p81-88, DOI: 10.1145/3051457.3051466


## Références logicielles


(Gradescope) http://gradescope.com
(INGInious) UCLouvain. INGInious. https://www.inginious.org
(Moodle) http://moodle.org