

Une démarche structurée pour définir quelques conseils limitant l'impact de la pandémie sur l'évaluation de nos étudiants

Pascal Detroz – p.detroz@uliege.be

Université de Liège – SMART-IFRES – DIDACTIfen

Laura Malay – l.malay@uliege.be

Université de Liège – SMART-IFRES

Vinciane Crahay – v.crahay@uliege.be

Université de Liège – SMART-IFRES

Pour citer cet article : Detroz, P, Malay, L, & Crahay, V. (2020). Une démarche structurée pour définir quelques conseils en vue de limiter l'impact de la pandémie sur l'évaluation de nos étudiants, Numéro Hors-série, 1, 97-110.

Résumé

La pandémie que nous vivons est tout à fait inédite. Elle a provoqué un basculement de nos enseignements qui se sont, dès lors, donnés à distance. Elle pose également des défis gigantesques concernant l'évaluation. En effet, nos dispositifs n'ont, pour la plupart, pas été initialement pensés pour être administrés à distance. Le recours à un **Modèle Structuré de Conception d'Évaluation des Étudiants (MoSCodEE)** dans l'enseignement supérieur, pour analyser l'impact de la pandémie sur notre manière d'évaluer, a l'avantage de guider notre regard sur différents moments de l'évaluation. De cette manière nous pourrions poser un regard holistique et systémique sur l'effet potentiellement délétère de la pandémie.

Mots-clés

Évaluation des étudiants, Conception d'évaluation, Pandémie.

Abstract

The pandemic we are experiencing now is completely unheard-of. It has forced us to move our classes online. It also radically challenges our assessment methods which have not been initially thought out to take place at a distance. Resorting to the **Structured Model of Student Assessment Design** in Higher Education in order to analyse the impact of the pandemic on our assessment methods makes it possible to direct our attention at various stages of the assessment process. By adopting the model, we shall be able to look holistically and systematically at the potential harmful effects of the pandemic on student assessment.

Keywords

Student assessment, Assessment design, Pandemic.

1. Introduction

Dans l'enseignement supérieur, en Fédération Wallonie-Bruxelles de Belgique, le mot d'ordre donné aux enseignants lors de cette pandémie est : les évaluations certificatives sont maintenues à distance. On peut s'inquiéter de cette mesure. En temps normaux, les évaluations certificatives dans l'enseignement universitaire sont déjà en tension (Detroz et al. 2017). Entre auditoires surpeuplés contraignant les enseignants à recourir - abusivement parfois - au testing standardisé et examens oraux où les étudiants se suivent à la chaîne devant un enseignant éreinté dont le jugement sera biaisé par de nombreux éléments mis en lumière il y a parfois un demi-siècle par la docimologie critique (par exemple, Bonniol & Piolat, 1971 ; Felouzis, 1993 ; Leclercq, Nicaise & Demeuse, 2004 ; Piéron, 1963), la situation de l'évaluation certificative est loin d'être optimale à ce niveau d'étude. Non pas que le testing standardisé ou les examens oraux soient de facto de mauvaises modalités d'évaluation – certains sont d'un très haut niveau – mais la massification des étudiants conjuguée au peu de formation des enseignants à l'évaluation nous amène souvent à nous interroger sur la qualité de la note dans l'enseignement supérieur. Nous le rappelons souvent lorsque nous formons des enseignants universitaires : si les étudiants connaissaient les principales études scientifiques sur la difficulté de noter les examens de manière équitable à l'université, ils seraient beaucoup plus vindicatifs quant aux dispositifs qui leur sont imposés.

Dans ce contexte inédit de pandémie, l'évaluation subit une profonde mutation qui en modifie les normes. Les examens en présentiel ne pourront être organisés. Les enseignants et les étudiants vont être contraints à des évaluations à distance, souvent pour la première fois de leur vie. Nous pouvons dès lors suspecter que la situation, déjà tendue en temps normal, va se dégrader fortement. D'une part, les matières n'ont pas été vues dans les mêmes conditions (moins d'explications, moins de possibilités d'éclaircissements -et par les professeurs et par les pairs en confinement aussi-...). D'autre part, les enseignants vont devoir, dans l'urgence, « innover » en termes d'évaluation et/ou changer de modalité(s), le plus souvent de manière contrainte et forcée.

Ils devront également calibrer de nouveaux outils alors qu'ils méconnaissent leur subtilité. Ils seront tenus à certains choix et surtout à certaines limites. De facto, une question fondamentale se posera à eux : « Dois-je utiliser des outils de contrôle antifraude même si je sais que ceux-ci vont augmenter l'anxiété des étudiants ? ». Cette anxiété ne va-t-elle pas altérer la qualité de la note ? Et si oui, puis-je me permettre de ne pas mettre en place ce type d'outil ? Si la fraude s'organise chez mes étudiants, l'intégrité de la note sera elle aussi influencée négativement. Comment trancher ce nœud gordien ?

Aucun choix ne sera facile et de très nombreuses questions vont se poser aux enseignants. L'objectif de cette contribution est d'utiliser le **Modèle Structuré de Conception d'Évaluation des Étudiants (MoSCodEE)** dans l'enseignement supérieur afin de présenter de manière structurée quelques difficultés auxquelles vont être soumis les enseignants. Et, évidemment, de structurer les conseils que nous pouvons leur offrir.

2. Le Modèle Structuré de Conception d'Évaluation des Étudiants (MoSCodEE)

Ce modèle est très largement inspiré des travaux de Jean-Luc Gilles et du modèle *Cycle of Construction and Quality Control in Standardized Tests - CCQCST* (Gilles et Leclercq, 1995;

Gilles, 2002, Gilles et al. 2005). Ce modèle comprend huit étapes, présentées sous forme de cycle, qui sont [1] Analysis, [2] Design, [3] Questions, [4] Training, [5] Testing, [6] Correction [7] Feedback, et [8] Régulation. Nous proposons d'ajouter à ce cycle, à la valeur heuristique certaine, une neuvième étape intercalée entre la première et la seconde que nous nommons « finalités et valeurs de l'évaluation ». Cette nouvelle étape nous semble fondamentale. D'après nos observations, les dispositifs d'évaluation mis en place par les enseignants dépendent très clairement des intentions visées à travers l'évaluation. Souhaitent-ils concevoir une évaluation formative ou une évaluation certificative ? Une évaluation normative ou une évaluation critériée ? Les évaluateurs vont-ils privilégier la couverture de l'ensemble de la matière ou d'un chapitre en particulier ? Autant de questions qui vont, selon nous, orienter leur choix dans la suite des itérations de ce modèle. Nous proposons aussi une autre révision du modèle original qui consiste à ne plus le limiter aux formes standardisées d'évaluation (il a fait ses preuves, selon nos observations pour tous les types d'évaluation dans l'enseignement supérieur) et proposons, dès lors, une nouvelle formulation pour certaines de ses étapes afin d'en élargir le spectre.

Ainsi, l'étape [Training] sera rebaptisée « Communication vers l'étudiant ». Il nous semble effectivement que si l'entraînement aux procédures d'examen est un moment important, il ne peut pas faire abstraction d'une vision plus globale de la communication vers l'étudiant qui devrait cibler, par exemple, les enjeux de l'évaluation, la matière faisant objet de l'examen, les types de modalités d'évaluation utilisées, le mode de correction...

Dans la même logique, d'une vision plus globale de l'évaluation, l'étape de [Testing] sera renommée « Prise de données », celle de [Questions] sera renommée « Paramétrage et conception du test »... Nous avons également francisé certains termes. Au final, nous proposons donc une vision revisitée du modèle – *CCQCST* (figure 1)

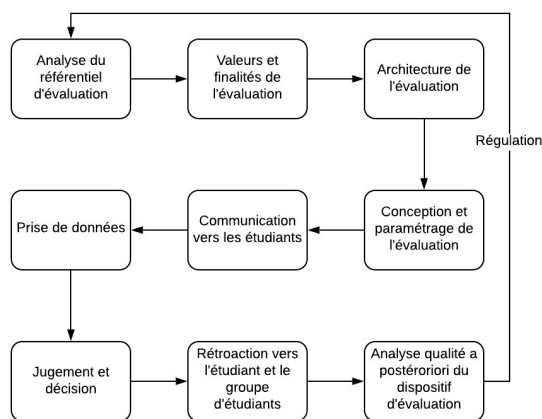


Figure 1. Adaptation du CCQCST de Gilles (Gilles et Leclercq, 1995 ; Gilles, 2002, Gilles *et al.* 2005)

Certes, il s'agit d'un processus présenté théoriquement comme chronologique et successif. Pour l'avoir testé dans nos accompagnements docimologiques avec les enseignants de notre institution, nous pouvons affirmer qu'il possède une réelle valeur heuristique : mettre la focale sur différentes parties d'une structure complexe. Il s'agit selon nous d'un « cadre dialogique », qui, selon notre expérience, peut permettre aux professionnels de développer

une activité nouvelle sur leurs activités antérieures, par un travail sur leur propre métier (Vanhulle, 2009).

Dans la suite de cette contribution, nous détaillerons les étapes de ce cycle de qualité et leur spécificité dans le cadre d'une évaluation urgente, non initialement planifiée et à distance, telle qu'imposée par notre contexte sanitaire. Nous proposerons aussi quelques conseils que nous avons donnés aux enseignants de notre institution. Ces conseils sont avant tout praxéologiques au sens de Vander Maren (2003). C'est-à-dire qu'ils puisent leurs racines dans une littérature scientifique maîtrisée par les auteurs mais qu'ils visent à s'adapter au contexte particulier des enseignants universitaires que nous accompagnons. Cette contribution se veut donc pragmatique plus que scientifique, ce qui est un mode de communication inhabituel pour nous mais qui nous semble pertinent et profitable dans ce contexte d'urgence.

3. Le modèle structuré de conception d'évaluation des étudiants dans l'enseignement supérieur

3.1. L'analyse du référentiel d'évaluation

Traditionnellement, une évaluation de qualité s'assure que l'alignement pédagogique (Biggs, 1996, Detroz, Tessaro & Younès, 2020) ait été pensé de manière efficace, ce que l'on peut résumer par la question : « Dans quelle mesure suis-je certain que les compétences que j'évalue ont fait l'objet d'un enseignement ? » Pour ce faire, la réalisation d'une table de spécification pour guider le processus d'évaluation est recommandée. Cet outil permet, dès lors, de renforcer l'alignement pédagogique nécessaire à toute évaluation.

Content Outline	Connaissance	Compréhension	Application	Analyse	Synthèse	Savoir-faire	Savoir etre
I. Basic Parts of Cell	1	2	0	0	0	0	0
A. Nucleus							
B. Cytoplasm							
C. Cell Membrane							
II. Plant vs Animal Cells	3	2	0	0	0	0	0
A. Similarities							
B. Differences							
1. Cell wall vs membrane							
2. food manufacture							
III. Cell Membrane	3	1	1	0	0	0	0
A. Living nature of							
B. Diffusion							
C. Substances diffused by cells							
IV. Division of Cells	4	3	2	4	1	4	0
A. Phases in Division							
B. Chromosomes and DNA							
C. Plant vs animal cell division							

Figure 2. Un exemple de table de spécification

Si l'on observe la table de spécification (Figure 2), dans la première colonne, l'enseignant est invité à lister les éléments qui ont été enseignés. Ces éléments sont regroupés et structurés en chapitres et sous-chapitres. Les colonnes suivantes portent sur les catégories de performance que l'enseignant souhaite voir atteintes par ses étudiants. Au croisement entre les éléments enseignés et les catégories de performance, est porté un chiffre qui correspond à la priorité que revêt cet apprentissage pour l'enseignant : dans cet exemple, de 0 à 4, 0 étant de très faible importance et 4 étant d'une importance majeure. En général, les priorités sont établies sur base de l'utilité qu'aura ce concept dans le parcours scolaire (s'il s'agit d'un prérequis) ou du parcours professionnel de l'étudiant.

L'utilisation d'une table de spécification est une démarche tout à fait classique en évaluation. Toutefois, la situation actuelle de pandémie amène quelques éléments de réflexions. Il est probable, par exemple, que le Covid 19 ait eu une influence sur le référentiel d'enseignement. Non pas sur le référentiel prescrit mais bien sur le référentiel réalisé. Par exemple, certains éléments de la matière n'ont peut-être pas été enseignés. Ou encore les méthodes d'enseignement ont pu être altérées par le passage à distance. Ainsi, si un cours dans lequel le débat occupait une place centrale a été remplacé par des podcasts essentiellement transmissifs, cela a un impact sur le niveau de performance atteint par ledit cours. Il en va de même si des TP ont été supprimés ou si des séances de répétition ont été annulées.

Or, si le référentiel d'enseignement a été modifié, le référentiel de l'évaluation doit s'adapter à cette réalité. Il s'agira donc d'ajuster la liste des contenus réellement enseignés, les catégories de performance atteintes ou les décisions en termes de priorités.

3.2. Valeurs et finalités de l'évaluation

C'est à cette étape que l'enseignant réfléchit aux modèles de l'évaluation dans lequel il souhaite s'inscrire (Bonniol & Vial, 1997, Vial, 2012). Évaluation certificative, évaluation formative, évaluation située authentique, évaluation gestion ou porteuse de sens, assessment for ou as learning, sont autant d'options qui s'offrent à lui et qui peuvent orienter certains choix dans les étapes ultérieures.

La situation inédite que nous vivons implique une grande réflexion à ce niveau. Que visons-nous à travers notre évaluation ? Nous l'avons écrit en introduction : les évaluations certificatives doivent être maintenues en Fédération Wallonie-Bruxelles. Mais dans le contexte particulier de pandémie, la bienveillance vis-à-vis des étudiants nous semble devoir être le principal mot d'ordre. Comment concilier ces deux finalités ? Maintenir l'intégrité des évaluations en utilisant des outils de proctoring¹ peut s'avérer à première vue utile au niveau de la certification mais cela s'avère terriblement anxiogène pour les étudiants. Or l'anxiété des étudiants est un facteur qui, lui aussi, peut nuire à l'intégrité de la note. Un étudiant trop anxieux ne sera sans doute pas capable de démontrer toute l'étendue de sa compétence. Augmenter le contrôle pour empêcher une minorité de frauder aura donc des effets délétères sur les étudiants les plus anxieux. Équilibrer cette équation est donc loin d'être aisé. Cela doit être réfléchi, décidé et assumé.

3.3. Architecture de l'évaluation

C'est à cette étape que le choix des modalités d'évaluation intervient. Ce choix se fait en cohérence avec l'analyse effectuée à l'étape « 1. Analyse du référentiel d'évaluation » et des finalités de l'évaluation définie à l'étape 2.

Habituellement dans l'enseignement supérieur, dans le cadre d'une évaluation certificative, nous conseillons aux enseignants de multiplier les modalités d'évaluation, les défauts des unes étant comblés par les qualités des autres. Nous leur conseillons ainsi de standardiser (Vrai-Faux, QCM...) l'évaluation portant sur des performances simples du type connaissance, compréhension ou application. Cela leur permet de préserver leur charge cognitive (et leur

¹ Proctoring est le terme consacré pour définir les méthodes et outils diminuant la fraude lors des examens à distance. Par exemple, filmer et analyser le comportement d'un étudiant lorsqu'il répond à un examen écrit.

précieus temps) qu'ils pourront alors mobiliser entièrement dans la correction de tâches qui nécessitent leur jugement évaluatif : la correction de productions longues visant des performances complexes.

Si l'on prend en compte le modèle SAMR (Puentendura, 2013) et qu'on l'adapte à l'évaluation à distance, deux catégories de choix sont possibles.

La première consiste à substituer à notre évaluation classique en présentiel des moyens technologiques pour qu'elle se fasse à distance dans des conditions les plus similaires possible. Cette solution a l'avantage de bousculer le moins possible les habitudes des étudiants et des enseignants en termes d'évaluation. Cependant, elle nécessite des adaptations si l'on veut s'assurer de procédures antifraude efficaces², procédures qui pourraient, par ailleurs, en fonction des options choisies, ne pas être sans effets sur la relation pédagogique que l'on souhaite entretenir avec les étudiants. Mesure-t-on, par exemple, les enjeux liés au renforcement de mesures démontrant une forme de méfiance ostensible vis-à-vis des étudiants ? D'autant plus que certaines mesures sont parfois relativement invasives (comme, par exemple, demander à l'étudiant de filmer à 360° le local où il se trouve lorsqu'il répond aux questions) et posent des questions juridiques et éthiques ?

La seconde catégorie consiste à transformer nos logiques d'évaluation pour les adapter à cette situation inédite. En effet, les mesures antifraude sont surtout nécessaires lorsque nous évaluons des performances simples du type connaissance. Si l'on demande à l'étudiant de donner la formule de la photosynthèse (question de connaissance), le recours à ses notes de cours ou l'aide d'un ami à distance (en vidéoconférence sur le téléphone, par exemple) sont des aides précieuses et illicites auxquelles il lui sera facile de recourir. Par contre, si on lui demande 5 points communs entre la vision de l'empire de Jules César et celle de Napoléon (question d'analyse), nous pouvons éventuellement lui permettre de consulter ses notes, limitant ainsi le recours à la fraude et diminuant les enjeux à contrôler son comportement de façon invasive. Si, pour évaluer ces tâches complexes, on demande par ailleurs à l'étudiant un travail personnel, spécifique et de volume significatif, les procédures classiques d'antiplagiat seront, dans la plupart des cas, suffisantes.

En conséquence, peut-être est-il envisageable de transformer une menace en opportunité, en modifiant nos évaluations de manière à renforcer leur caractère authentique, situé et contextualisé qui sont des caractéristiques connues pour renforcer la qualité en évaluation. Cette piste nous semble en tout cas prometteuse et à creuser, même s'il est probable que les conditions harmonieuses de cette transformation ne sont pas réunies : l'urgence se prête mal à la maturation d'un tel changement de paradigme.

Dans ce contexte précipité d'évaluation à distance, et afin de soutenir les enseignants dans le choix de modalités d'évaluation les plus judicieuses à mettre en place, nous leur avons asséné trois mots d'ordre : [1] Recyclez [2] Keep calm ! [3] Préservez la charge cognitive (la vôtre et celle de vos étudiants).

Le mot d'ordre « recyclez » tenait au fait que certains enseignants disposaient déjà de matériaux évaluatifs collationnés avant la pandémie dans de bonnes conditions (tests intermédiaires, tests formatifs, résultats aux TP...). L'idée sous-jacente étant de réfléchir à la

² En effet, à distance il n'est pas certain que l'étudiant connecté soit bien celui qui doit être évalué, ni qu'il le fera sans recourir à ses notes, par exemple, ou à différentes stratégies de « triche ».

possibilité de recycler ce matériau en vue d'une évaluation certificative finale, quitte à ouvrir un dialogue collaboratif avec les étudiants concernés.

L'idée de rester calme nous semblait utile. La plupart des Learning Management System (Moodle, par exemple) ou des outils d'évaluation sont très créatifs en termes de modalités de questionnement : appariements, mots croisés, textes à trous... Ces outils sont intéressants lors des évaluations formatives car ils leur amènent notamment un caractère ludique. Dans le cadre d'évaluations certificatives de fin d'année, il n'est pas certain qu'ils soient suffisamment robustes et pertinents ; il semble donc important d'analyser les options de ces outils pour s'en assurer et peut-être, le cas échéant, de rester fidèle à des modalités plus classiques et bien maîtrisées par les enseignants et les étudiants. Nous voulions aussi insister, via ce conseil, sur la nécessité de bien mesurer et anticiper l'impact d'éventuels outils numériques sur la performance des étudiants : il était sans doute sage de recourir, si possible, à des modalités d'évaluation nécessitant peu d'appropriations technologiques chez les étudiants.

Nous avons enfin recommandé aux enseignants de réfléchir à la charge cognitive des étudiants, mais aussi à la leur. Nous leur avons proposé ainsi de prévoir de courtes séances d'évaluation entrecoupées de pauses. Par exemple, organiser un QCM d'une heure, octroyer 30 minutes de pause avant de fournir aux étudiants les questions à réponses ouvertes longues. En cas d'examen oral, nous leur avons suggéré de prévoir de courtes pauses entre étudiants ou groupes d'étudiants pour leur offrir une respiration. Interroger à distance nous semble effectivement plus éreintant encore qu'interroger en présentiel.

3.4. Conception et paramétrage de l'évaluation

C'est à cette étape que va être opérationnalisée et peaufinée l'architecture qui a été réfléchie à l'étape précédente. Les questions vont y être rédigées, le test paramétré, les barèmes de points choisis, les pondérations entre questions décidées...

Exploiter la table de spécification (cf. « 1. Analyse du référentiel d'évaluation) pour piloter la conception des questions permettra de s'assurer que le dispositif d'évaluation couvre bien les éléments enseignés et les processus mentaux les plus importants parmi ceux abordés en cours.

Différents contrôles de la qualité des items rédigés devront être réalisés. La littérature offre des conseils pour la rédaction de divers types de questions (Voir le manuel de Brookhart & Nitko, 2019, par exemple). Le recours à la relecture des items par des experts (par exemple, les assistants du cours dans l'enseignement universitaire) est également à recommander.

Certains conseils varient en fonction des modalités de questionnement utilisées.

Dans le cadre du testing standardisé de type QCM ou Vrai-Faux, il est judicieux de renvoyer classiquement vers les règles de rédaction qui assureront une bonne qualité psychométrique des items rédigés (Leclercq, 1986). À distance, il conviendra de réfléchir à la mise en place de procédures antifraude plus affirmées. En présentiel, des formes parallèles de l'examen sur la base d'un mélange des questions suffit amplement à diminuer la possibilité de fraude... À distance, au vu des possibilités de triche comme le partage d'écran par exemple, d'autres possibilités s'offrent aux concepteurs comme mélanger les questions, mélanger les

propositions, laisser un temps limité imparti par question (ou pour l'ensemble du test)³, mettre éventuellement en place des outils ou démarches de *proctoring* qui assureront que l'étudiant ne bénéficiera pas d'aide illicite pour répondre aux questions.

Ces mesures n'élimineront toutefois pas entièrement les possibilités de triche mais les limiteront et les rendront plus complexes. Certaines d'entre elles généreront beaucoup d'anxiété chez les étudiants.

Une autre option de passation en ligne doit être discutée : l'attribution, à chaque étudiant, d'un choix aléatoire de questions puisées dans une banque d'items. Chaque étudiant répondra donc à un set différent de questions. Habituellement, nous rejetons cette option pour deux raisons : [1] les questions étant tirées au hasard, la procédure ne garantit pas la validité de contenu et de construit (cf. « 1. Analyse du référentiel d'évaluation ») et [2] elle n'est pas équitable car elle ne garantit pas le même niveau de difficulté pour tous les étudiants⁴. Par contre, en temps de pandémie, nous sommes moins assertifs car il peut s'agir là d'une procédure antifraude efficace. Certains logiciels permettent de catégoriser les questions selon leur niveau de difficulté, le chapitre auquel elles se réfèrent et le processus mental qu'elles visent. C'est intéressant car ces options permettent d'effectuer, pour chaque étudiant, un échantillonnage stratifié de questions garantissant, par exemple, qu'ils aient le même nombre de questions pour chacun des chapitres.

Dans le cadre des questions à réponses ouvertes longues, certaines procédures antifraude restent pertinentes comme le mélange de questions. De manière plus spécifique à cette modalité, une bonne façon de procéder est d'éviter de poser des questions de connaissance pour favoriser des questions de niveau taxonomique supérieur qui demandent un développement long, voire nécessitent une prise de position personnelle. On veillera toutefois à préserver l'alignement pédagogique. Il sera également nécessaire de recourir à des outils anti-plagiat pour analyser les similitudes éventuelles entre réponses.

Pour les productions longues (travaux, portfolios, cartes conceptuelles ...) asynchrones, il est recommandé de mettre beaucoup de soin et de rigueur pour préciser les consignes. Dans ce contexte de pandémie, les étudiants éprouvent sans doute plus de difficultés pour obtenir de l'information informelle auprès des enseignants ou des pairs. Tous les détails doivent donc être explicités. Par ailleurs, prévoir, dans la structuration du travail, un moment pour l'expression personnelle des étudiants quant à la matière est également intéressant car ce travail personnel ne peut être que difficilement plagié. La méthode Pecha Kucha nous semble intéressante de ce point de vue. A ce titre, nous vous renvoyons vers le texte de Christelle Lison dans le présent numéro. Évidemment, pour les travaux, l'usage des outils anti-plagiat est également recommandé.

L'« examen oral » quant à lui est la modalité qui semble peut-être de prime abord la plus « facile » à transposer à distance. Il s'agit toutefois aussi de la modalité pour laquelle on retrouve, en présentiel, le plus de biais de correction et de biais liés à la distribution des questions (au hasard ou distribuées par l'enseignant, elles ne couvrent en général pas toute la matière). Organiser un examen oral à distance est inédit pour les enseignants et les étudiants.

³ Si cette mesure est intéressante dans le cadre de procédures antifraude, elle est toutefois génératrice d'anxiété chez les étudiants.

⁴ Il est difficile, même pour un expert, de déterminer avec certitude le degré de difficulté d'une question par rapport à une autre. Il est possible de pallier cette difficulté en recourant au Modèle de Réponse à l'ITEM, mais cette solution nécessite de l'expertise que les enseignants ne peuvent acquérir dans un contexte d'urgence.

Leur stress habituel -à tous deux- sera assurément décuplé : les éventuels retards de l'enseignant seront invivables pour des étudiants qui, derrière leur écran, se demanderont pourquoi on ne les contacte pas (ils risquent bien de vérifier leur wifi toutes les deux secondes). Les étudiants ne se sentiront probablement pas à l'aise face à cet écran qui servira d'interface et les plongera dans une ambiance très lointaine d'une situation d'échanges humains. Ils pourront par ailleurs enregistrer la séance avec leur GSM : l'équité et le comportement de l'enseignant deviendront alors facilement publics ou publiables⁵. Ajoutons que les questions de l'enseignant vont circuler plus vite mais aussi les réponses ou pistes de réponses que l'enseignant aura données pour mettre l'étudiant sur la voie. En conséquence, cela favorisera la fraude par ceux qui seront sur les « bons » réseaux. L'enseignant va être plus stressé -lui aussi- face à la situation d'évaluation inhabituelle et sa charge cognitive sera beaucoup plus importante, ce qui va augmenter les biais de jugement déjà importants à la base.

Nous risquons bien, compte tenu de tout cela, de récompenser des étudiants ayant mis en place les stratégies les plus efficaces (y compris la fraude) plutôt que ceux ayant une maîtrise attendue de la matière mais des stratégies moins porteuses dans ce contexte d'évaluation à distance.

3.5. Communication auprès des étudiants

Cette étape est celle où l'enseignant communique les informations utiles aux étudiants en lien avec les évaluations. Il leur transmet, par exemple, la table de spécification (ou en tout cas la matière sur laquelle va porter l'évaluation) et les modalités d'évaluation. Idéalement, il leur propose également des tests d'entraînement, leur permettant ainsi de se familiariser avec la procédure d'évaluation. Ces tests fourniront par ailleurs à l'enseignant l'occasion de donner un dernier feedback aux étudiants.

Cette étape est importante car elle permet à l'étudiant de diminuer sa charge mentale et son stress lors de l'évaluation certificative, sachant qu'il a été préparé au dispositif qui va lui être proposé.

Organiser un test à blanc aura également des vertus en ce qui concerne les enseignants. Cela leur permettra, à eux aussi, de se familiariser avec l'outil et d'être confrontés une première fois à l'environnement d'évaluation et aux nombreux choix qu'il s'avèrera nécessaire de poser en termes de paramétrages. Cela permettra également de prévoir, sur base de cette expérience, des routines pour solutionner les problèmes les plus fréquents que les étudiants risquent de rencontrer (problèmes de connexion, perte temporaire de connexion internet, panne de batterie, « plantages » d'ordinateur durant l'examen...)

Dans tous les cas, même si un test à blanc est impossible ou difficile, il sera nécessaire de structurer l'information à destination des étudiants avec énormément de soin. Ils ne disposeront pas des canaux habituels (demander à un condisciple qu'ils croisent, poser la question lors de l'intercours ou à la fin de celui-ci...) qui assurent la redondance de l'information. Pour les amener à affiner leur compréhension des procédures évaluatives auxquelles ils vont être soumis, il faudra donc être précis et ne laisser aucune place à l'ambiguïté. Ainsi, dans notre institution, nous avons conseillé à nos enseignants de revoir leur plan de cours en précisant la matière qui sera abordée lors de l'examen, le niveau d'exigence pour chacun des éléments de cette matière, les modalités d'évaluation et leurs

⁵ Les réseaux sociaux seront sans concession avec les captures d'écran.

caractéristiques (temps imparti ou non, livre ouvert ou non, barèmes utilisés, matériaux à disposition...), les modalités pratiques de l'examen (date, heure, temps de passation, outils utilisés, modalités de communication...) et tout autre élément utile aux étudiants.

3.6. *Prise de données*

L'évaluation prend place à proprement parler à ce niveau : c'est le moment où l'étudiant démontre sa compétence à travers une performance qu'il réalise en réponse à un dispositif évaluatif mis à sa disposition.

Dans le cas d'un travail à réaliser par l'étudiant (travaux, cartes conceptuelles, portfolios...), la réalisation de cette étape n'est pas sous contrôle de l'enseignant : elle se réalise à domicile. Cette modalité sera peu impactée par la pandémie que nous connaissons.

Pour les examens écrits à distance, la présence sociale de l'enseignant est un facteur important pour réduire l'anxiété des étudiants. On peut, par exemple, imaginer que l'enseignant communique avec ses étudiants de manière synchrone, juste avant l'examen, pour leur donner les dernières consignes et les encouragements d'usage. Il devra aussi être disponible pendant l'examen, par courriel ou téléphone, pour pouvoir réagir à des soucis techniques rencontrés par les étudiants.

L'examen oral, déjà stressant pour certains étudiants en présentiel sera encore plus anxiogène à distance. En effet, au stress classique - lié à la maîtrise de la matière et à l'interaction avec l'enseignant - s'ajoutera le stress inédit lié à la gestion des outils technologiques et des éventuelles démarches de proctoring. Par ailleurs, certains étudiants seront confinés dans un espace restreint avec leurs proches ce qui provoquera une autre source de stress. Si cette modalité d'évaluation est choisie, une attitude extrêmement bienveillante sera nécessaire. Il ne s'agit pas seulement d'un conseil humaniste. Selon nous, il en va de l'intégrité de l'épreuve. En effet, le stress de l'étudiant pourrait inhiber l'accès à ses propres connaissances ce qui biaiserait l'information ainsi récoltée par l'enseignant. Il nous semble aussi nécessaire de questionner ce type d'évaluation au regard des inégalités sociales. Disposer d'un environnement calme et de matériel performant est l'apanage d'une catégorie spécifique d'étudiants, sans doute les plus favorisés.

Pour préserver leur capacité de bienveillance, nous avons conseillé aux enseignants de se ménager des pauses de manière plus fréquente qu'en présentiel. La fatigue résultant d'une interaction en ligne nous semble en effet plus importante que celle que l'on pourrait avoir en face à face. Se ménager des plages de repos s'avèrera d'autant plus nécessaire.

Une réponse appropriée à un enseignement qui s'organise à distance consiste à effectuer des évaluations continues, tout au long de l'année. Demander aux étudiants de compléter une carte conceptuelle en fonction de ses lectures, tracer l'activité des étudiants sur les outils numériques, leur proposer des quizz fréquents... sont des activités qui permettent d'organiser le temps d'étude et d'appropriation de la matière des étudiants. Cela permet également d'obtenir des mesures répétées utiles à une évaluation certificative et de donner des feedbacks fréquents aux étudiants quant à leur évolution. Avouons toutefois que, dans ce contexte de pandémie, où le basculement de nos enseignements en présentiel s'est fait dans l'urgence, les conditions d'un enseignement à distance de qualité n'ont pas pu être réunies (Hodges et al., 2020). Cela constat s'adresse aussi à la mise en place d'évaluations continues.

3.7. Jugement et décision

Suite à la récolte des données, une étape de jugement et de décision est nécessaire. Elle aboutira à l'attribution d'une note à l'étudiant. Les conseils que nous donnons aux enseignants hors pandémie restent, selon nous, d'actualité.

Si nos évaluations sont de type standardisé (QCM, Vrai-Faux), nous recommandons d'analyser la qualité des questions à posteriori, par exemple en utilisant la corrélation bisériale de point. La plupart des outils d'évaluation en ligne n'offrent cependant pas cette option, ou pas avec les outils adéquats. C'est un sujet qui doit sans doute mobiliser les équipes informatiques avant la session.

Pour ce qui concerne les corrections de questions à réponses ouvertes longues, les rapports, les travaux, les cartes conceptuelles, les conseils habituels restent de mise. Une grille descriptive d'évaluation, que l'on prendra la peine de calibrer, s'avère toujours autant nécessaire.

Si l'enseignant décide de recourir à un examen oral à distance, il peut être utile de garder une trace de l'évaluation, par exemple en enregistrant l'examen. Cela permettra d'effectuer quelques contrôles qualité quant à la note attribuée à chaque étudiant. Nous l'écrivions en introduction, la docimologie⁶ critique a répertorié de nombreux biais dans nos dispositifs de correction, mettant en évidence l'importance des tels contrôles qualité sur nos procédures. Par exemple, peut-on être certain qu'un étudiant évalué en début de session aurait eu les mêmes points s'il avait été évalué en fin de session ? La sévérité de l'enseignant a-t-elle évolué au fil des passations ? Pour répondre à ces questionnements légitimes, revoir certains enregistrements avant de confirmer définitivement les notes attribuées peut avoir un réel intérêt heuristique.

3.8. Rétroaction vers l'étudiant et le groupe d'étudiant

Une fois l'épreuve validée et entérinée, la mise à disposition de feedbacks — de rétro-informations — pour les étudiants leur permet d'avoir un retour sur l'épreuve qu'ils ont passée et, ainsi, d'avoir un aperçu de leurs points forts et de leurs faiblesses. Ce sont des informations très utiles en vue notamment d'une préparation (plus) efficace à la seconde session. Nous n'avons pas l'espace, ici, pour détailler l'importance des feedbacks dans l'enseignement, ni de définir ce que sont des feedbacks de qualité. Nous vous renvoyons sur la littérature abondante sur le sujet (par exemple, Nicol et Macfarlane-Dick, 2006 ; Hattie & Timperley, 2007 ; Sadler, 2010 ; Narciss, 2013 ; Winstone & Carless, 2019).

Il y a des enjeux importants à donner des feedbacks de qualité de manière rapide aux étudiants. Certaines plateformes permettent aux étudiants de monitorer leur test de type QCM en leur donnant immédiatement la réponse correcte à la question à laquelle ils viennent de répondre et en incrémentant un score en continu durant la passation. Nous ne recommandons pas cette option et ce pour trois raisons principales :

- Comme nous l'avons signalé à l'étape 6, nous recommandons que le test fasse l'objet d'une procédure qualité avant la diffusion des résultats ;
- Avoir une vision en temps réel de son score peut être stressant pour l'étudiant. Ce stress peut être inhibant chez l'étudiant. Par ailleurs, il peut également l'amener à

⁶ Discipline scientifique consacrée à l'étude du déroulement des évaluations en pédagogie.

prendre des décisions contre-productives (répondre en prenant des risques dans un contexte d'incertitude élevée, par exemple) ;

- Cela augmente le risque de fraude, par partage des réponses correctes reçues.

Par contre, un feedback complet comprenant au moins les réponses correctes attendues ou un corrigé type est fortement conseillé, et ce le plus rapidement possible après le test.

Dans ce contexte particulier, il semble pertinent également d'expliquer aux étudiants les éventuels contrôles qualité qui ont pu être faits sur les résultats, en tout cas lorsqu'ils sont à l'avantage des étudiants. Par exemple, certains enseignants de notre institution vont ajuster la moyenne des étudiants à celles des étudiants des années précédentes s'ils constatent un taux d'échec plus important lors des évaluations.

3.9. Analyse qualité a posteriori du dispositif d'évaluation

Cette étape permet de récolter de l'information sur le processus d'évaluation mis en œuvre et de procéder à des réajustements éventuels pour les examens futurs. Bien évidemment, on peut espérer que l'épisode que nous vivons est temporaire et que le cours d'une vie universitaire plus classique s'en suivra. Il nous semble utile de prévoir une analyse à froid des procédures d'examens que nous avons mises en œuvre dans ce contexte de crise.

Plus fondamentalement, face aux difficultés que nous éprouvons, se poser la question de la pertinence d'une évaluation avant tout sanctionnante prenant place exclusivement en fin d'année semble raisonnable. L'évaluation continue mérite sans doute un intérêt un peu plus prononcé. C'est peut-être l'occasion d'amener ce débat dans nos universités.

4. Conclusions

Nous avons pris plaisir à rassembler et structurer quelques conseils et à les communiquer via ce numéro spécial. Nous aimerions rappeler, dans cette conclusion, la portée de ce texte. Nous l'avons voulu pragmatique, offrant quelques pistes d'action ou de réflexion aux enseignants universitaires qui nous feraient le plaisir de le lire. Certes, nos conseils auraient pu être étayés plus systématiquement par des textes issus de littérature. Cela lui aurait donné une forme plus conventionnelle. Mais, d'une part, nous n'aurions pas pu finaliser dans un timing contraint tel que celui de ce numéro spécial. D'autre part, et ce n'est pas un détail, cela aurait sans doute nui à sa lisibilité. Nous nous rendons compte que le parti pris que nous avons choisi est discutable. Il nous semble toutefois s'inscrire assez bien dans la logique qui a présidé à l'émergence de ce numéro spécial. Logique qui vise à ce que les acteurs de l'évaluation, quel que soit leur niveau de conceptualisation de ce domaine, puissent bénéficier de conseils utiles. C'est sans doute, au final, l'intérêt de ce travail. Dans notre Institution, nous avons été très agréablement surpris de voir l'abnégation des enseignants avec lesquels nous travaillons et leur souci d'atteindre un niveau de qualité en évaluation malgré un contexte pandémique qui s'y prête peu. Nous avons été heureux, aussi, de voir que des inquiétudes bien légitimes ont donné lieu à une mobilisation importante et à un niveau élevé d'entraider entre collègues. Si ce texte permet de soutenir de tels élans, alors, il aura atteint sa cible.

5. Bibliographie

- Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher education*, 32(3), 347-364. <https://doi.org/10.1007/BF00138871>
- Bonniol, J., & Piolat, M. (1971). *Correction en mathématique et en anglais*. Actes du XVII^e Congrès international de psychologie appliquée, 1179-1189.
- Brookhart, S., & Nitko, A. (2019). *Educational Assessment of Students*. (8th ed.). Pearson.
- Detroz, P., Romainville, M., Auquièrre, A., Dony, M., Piazza, A., & Vause, A. (2017). *Comprendre et améliorer les pratiques d'évaluation des acquis des étudiants dans l'enseignement supérieur*. Rapport d'expertise. Fédération Wallonie Bruxelles. <http://hdl.handle.net/2268/206324>
- Detroz, P., Tessaro, W., & Younès N. (2020). Pour la relance d'une évaluation congruente à l'université. *Évaluer. Journal international de recherche en éducation et formation*, Numéro Hors-série, 1.
- Felouzis, F. (1993). Interactions en classe et réussite scolaire : Une analyse des différences filles-garçons. *Revue française de sociologie*, 34(2), 199-222. <https://doi.org/10.2307/3322488>
- Gilles, J.-L. (2002). *Qualité spectrale des tests standardisés universitaires*. Thèse de doctorat. Université de Liège, Belgique.
- Gilles, J.-L., & Leclercq, D. (1995). *Procédures d'évaluation adaptées à des grands groupes d'étudiants universitaires - Enjeux et solutions pratiquées à la FAPSE-ULG*. Actes du Symposium International sur la Rénovation Didactique en Biologie. Tunis : Université de Tunis.
- Gilles, J.-L., Piette, S.-A., Detroz, P., Tinnirello, S., Pirson, M., Dabo, M., & Lê, H. (2005). The electronic Construction and Quality Control in Standardized Testing platform project (e-C&QCST). *European Association For Research on Learning and Instruction (EARLI) 11th Biennial Conference of EARLI*. Nicosia: University of Cyprus.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81-112.
- Hodges, C., Moore, S., Lockee, B., Trust, T., & Bond, A. (2020). The Difference Between Emergency Remote Teaching and Online Learning. *Educause review*. Retrouvé sur : <https://er.educause.edu/articles/2020/3/the-difference-between-emergency-remote-teaching-and-online-learning>
- Laurier, M. (1998). Bonniol, J.-J., & Vial, M. (1997). Les modèles de l'évaluation : textes fondateurs avec commentaires. Bruxelles : De Boeck Université. *Revue des sciences de l'éducation*, 24(3), 691-. <https://doi.org/10.7202/031989ar>
- Leclercq, D. (1986). *La conception des questions à choix multiple*. Labor.
- Leclercq, D., Nicaise, J., & Demeuse, M. (2004). Docimologie critique : des difficultés de noter des copies et d'attribuer des notes aux élèves. In M. Demeuse M. (Ed.), *Introduction aux théories et aux méthodes de la mesure en sciences psychologiques et en sciences de l'éducation* (pp. 273-292). Les éditions de l'Université de Liège.
- Narciss, S. (2013). Designing and Evaluating Tutoring Feedback Strategies for digital learning environments on the basis of the Interactive Tutoring Feedback Model. *Digital Education Review*, 23, 7-26- <http://greav.ub.edu/der/>
- Nicol, D.-J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: a model and seven principles of good feedback practice. *Studies in Higher Education*, 31, 199-218.
- Pieron, H. (1963). *Examens et docimologie*. Presses Universitaires de France.
- Puentedura, R. (2013). SAMR and TPACK: An Introduction. *Hipassus, Blogue de Ruben R. Puentedura*, 14.

- Sadler, D. (2010). Beyond feedback : Developing student capability in complex appraisal. *Assessment & Evaluation in Higher Education*, 35(5), 535-550. <https://doi.org/10.1080/02602930903541015>
- Van Der Maren, J.-M. (2003). *La recherche appliquée en pédagogie : Des modèles pour l'enseignement*. De Boeck Supérieur. doi:10.3917/dbu.maren.2003.01.
- Vanhulle, S. (2009). Évaluer la professionnalité émergente des futurs enseignants : un pari entre cadres contraignants et tensions formatives. In L. Mottier Lopez (Ed.), *Évaluations en tension : Entre la régulation des apprentissages et le pilotage des systèmes* (pp. 165-180). De Boeck Supérieur.
- Vial, M. (2012). *Se repérer dans les modèles de l'évaluation : méthodes, dispositifs, outils*. De Boeck.
- Winstone, N., & Carless, D. (2019). *Designing effective feedback processes in higher education: A learning-focused approach*. Routledge.