



**Les journées de
l'interdisciplinarité**

Les usages des tests non rédactionnels dans l'évaluation des connaissances à distance des étudiants de Pharmacie : impact du déploiement du e-learning lors de la crise sanitaire du COVID-19

The uses of non-editorial tests in the distance assessment of Pharmacy students: impact of the deployment of e-learning during the COVID-19 crisis

Claire NIKITPOULOS

Département DUENES – Université de Limoges

Nicolas PICARD

Département DUENES – Université de Limoges

URL : <https://www.unilim.fr/journees-interdisciplinarite/99>

DOI : 10.25965/lji.99

Licence : CC BY-NC-ND 4.0 International

Résumé : L'implémentation de tests non rédactionnels entraîne de nouveaux usages dans l'évaluation des connaissances des étudiants. Cette recherche s'intéresse à l'évolution des tests non-rédactionnels et à l'impact du niveau de maîtrise des enseignants sur Moodle dans le choix du format et du niveau de complexité des questions. Il ressort de notre étude que les formats de questions plus traditionnels restent les plus utilisés dans les évaluations formatives à distance. Cependant, la question cloze, plus récente et plus complexe dans sa mise en œuvre, a été fortement employée dans les évaluations de fin d'année afin d'obtenir un système de notation homogène entre les différentes disciplines. L'étude montre également que le choix des formats de questions employés dans les tests non rédactionnels est lié à d'autres facteurs que l'objectif pédagogique à atteindre.

Mots clés : tests non rédactionnels, évaluations, exercices interactifs, QCM, e-learning

Abstract: The implementation of non-editorial testing is leading to new uses in the assessment of student knowledge. This research is interested in the evolution of non-editorial tests and the impact of the level of mastery of teachers on Moodle in the choice of format and level of complexity of questions. Our study shows that the more traditional question formats are still the most used in distance formative assessments. However, the cloze question, newer and more complex in its implementation, has been widely used in end-of-year assessments in order to achieve a scoring system similar among the different disciplines. The study also shows that the choice of question formats used in non-editorial tests is linked to factors other than the educational objective to be achieved.

Keywords: non-editorial tests, evaluations, interactive exercises, MCQ, e-learning

Introduction

Depuis 1915, le recours aux tests non rédactionnels (QCM, QCU, vrai-faux, etc.) s'est généralisé dans le milieu de l'enseignement. Leur évolution est marquée par l'apparition progressive au XX^{ème} siècle de tests de standardisation de l'évaluation, de l'ordinateur puis du e-learning vers la fin de ce siècle. Ce dernier amène avec lui de nouvelles formes de questions (textes à trous, phrases à relier, éléments à glisser-déposer sur une image, etc.) qui enrichissent les formations. Mais c'est en 2020, pour faire face à la crise sanitaire liée à la COVID-19, que l'évaluation des connaissances à distance dans les formations universitaires s'est généralisée brutalement.

1. Revue de littérature

1.1. L'origine des QCM

En 1914, Frederick J. Kelly (1880-1959) soutient une thèse intitulée « Teachers' Marks : Their Variability and Standardization » à l'Université de Columbia. Ses travaux de recherche portent sur la notation des élèves qui lui semble trop subjective et le temps de correction des évaluations jugé excessif. Il propose alors une standardisation de l'évaluation. Devenu professeur à l'Université du Kansas, Kelly réalise en 1915 le « Kansas Silent Reading Test » pour évaluer rapidement les élèves de 8 à 18 ans. Les tests contiennent des QROC (Questions à Réponse Ouverte Courte), des QCU (Questions à Choix Unique) ainsi que des questions à choix numériques. Les élèves doivent ainsi écrire ou entourer les bonnes réponses ou barrer les mauvaises sur leurs copies.

On retrouve ensuite ce format d'évaluation dans le test « Army Alpha » de l'armée américaine pour mesurer rapidement les capacités des recrues pour la Première Guerre Mondiale. Plus d'un million et demi d'hommes ont réalisé cette épreuve mesurant les capacités intellectuelles et émotionnelles des soldats. L'Army Alpha comportait 8 tests verbaux : directions orales, problèmes arithmétiques, jugement pratique, synonymes-antonymes, phrases en désordre, séries de nombres à compléter, analogies, informations générales. Chaque test employait un format particulier : vrai-faux à souligner, réponse numérique, QCU à souligner, même/opposé à souligner, QCU à cocher, nombre à souligner. Les résultats déterminaient la capacité de service du soldat, sa classification d'emploi et son potentiel pour un poste de leadership. Les soldats analphabètes ou non anglophones, devaient passer le test « Army Beta », l'équivalent de l'examen en non verbal. Le succès des tests standardisés les mène à leur généralisation aux USA. En 1921, 2 millions de soldats passent leurs tests d'entrée à l'armée (Army Alpha ou Army Beta) et 3 millions d'élèves réalisent leurs épreuves sous ce format.

Vers le début des années 20, Benjamin Wood (1894-1986) psychologue à l'Université de Columbia, développe la question à choix multiples. Ce format contient une racine (question, problème à résoudre, déclaration incomplète à compléter, etc.) et des réponses possibles. L'apprenant peut choisir une ou plusieurs réponses. En 1926, le Stanford Achievement Test (SAT) ajoute aux autres formats de questions, les items à choix multiples pour évaluer les connaissances des élèves de l'*Elementary School* (équivalent du primaire) au *High School* (équivalent du lycée) aux USA. E.F Lindquist (1901-1978), Professeur d'Éducation à l'Université de l'Iowa, introduit dès 1959 l'ACT (American College Testing), un test standardisé utilisé pour évaluer les connaissances des élèves avant l'admission au collège.

Les facultés de Médecine et de Pharmacie Françaises voient l'arrivée du lecteur optique de marques en 1960. Ce dernier facilite la correction des copies des étudiants en les corrigeant automatiquement et rapidement. L'apparition des ordinateurs en 1970 entraîne une augmentation des usages en matière de tests non rédactionnels. Différentes expérimentations sont réalisées pour générer des examens à partir d'un ordinateur (Dudley, 1973 ; Cartwright et Derevensky, 1975, etc.). Dudley (1973) étudie l'impact de l'ordinateur sur la stimulation des élèves et sur l'aide que ce dernier peut apporter pour tester les progrès des élèves. L'ordinateur devint alors une solution pour fournir des procédures efficaces, peu coûteuses et de qualité pour des évaluations répétitives (Cartwright, 1975).

L'arrivée du e-learning dans le milieu des années 90 et des plateformes d'enseignement à distance (Moodle, Claroline, Blackboard, etc.) permet aux tests non rédactionnels de s'enrichir en nouveaux formats de questions (textes à trous, mots croisés, phrases à relier, éléments à glisser-déposer, etc.) et en nouvelles fonctionnalités (correction automatique, rétroaction automatisée, paramétrage manuel du système de notation, etc.).

1.2. Niveau de complexité et de difficulté de la question

Dans la construction du test, rédactionnel ou non rédactionnel, l'enseignant doit gérer le niveau de complexité et la difficulté de chaque question en fonction des objectifs pédagogiques. Pour cela, l'emploi de la taxonomie de B. Bloom (1956) est préconisé pour élever le niveau de réflexion chez l'apprenant. Ce modèle d'habiletés intellectuelles permet de définir le niveau d'apprentissage souhaitée et d'élaborer une séquence pédagogique appropriée aidant l'apprenant à atteindre l'objectif pédagogique. Bloom définit six niveaux de comportements intellectuels importants dans l'apprentissage, allant du plus simple au plus complexe :

- Le niveau « Connaissance » est défini comme le simple rappel d'un sujet appris précédemment. Cependant, il n'y a aucune certitude que l'apprenant ait compris le sujet.
- Le niveau « Compréhension » se réfère à la capacité de donner du sens au sujet appris, en la rendant disponible pour une utilisation future.
- Le niveau « Application » est la capacité à mettre en pratique ses acquis, à appliquer une méthode ou une théorie.
- Le niveau « Analyse » relève de la capacité à analyser, à distinguer, à déduire ou encore à comparer des informations.
- Le niveau « Synthèse » correspond à la capacité à produire une réflexion personnelle, à composer, à inférer ou encore à raisonner sur des informations.
- Le niveau « Evaluation » est la capacité à juger, à évaluer, à comparer, à discriminer et à argumenter sur un sujet.

Cependant, D.A. Sousa (2002), révèle une confusion limitant l'utilisation de la taxonomie de Bloom et l'augmentation du niveau de réflexion des apprenants :

« La complexité et la difficulté relèvent d'opérations mentales radicalement différentes, mais elles sont souvent interprétées comme étant synonymes. On les traite comme si l'on s'agissait d'une seule et même chose, comme si l'une se substituait à l'autre. » (D.A. Sousa, 2002, p. 280).

Selon lui, « la complexité décrit le processus de réflexion que le cerveau utilise pour traiter l'information » alors que la difficulté correspond « à l'intensité de l'effort que l'apprenant doit fournir à un niveau de complexité donné afin d'atteindre l'objectif d'apprentissage ». Ainsi, il est possible de monter le niveau de difficulté de la question sans pour autant modifier son niveau de complexité.

Figure 1. Niveaux de complexité et de difficulté



Les différents niveaux de complexité et de difficulté peuvent être employés en adaptant les questions et les propositions de réponses des tests non rédactionnels. En fonction de la question et de la ou des réponses attendues, les enseignants vont choisir le format de questions adapté. Les tests non rédactionnels peuvent mesurer les hauts niveaux de complexité (analyse, synthèse et évaluation) si les questions conduisent à considérer les connaissances de façon plus profonde (Baillifard et Thurre-Millius, 2020). Il est possible d'interroger des relations (temporelles, logiques, etc.) entre les concepts, de proposer des analogies pour forcer l'étudiant à comparer des notions ou encore des questions qui nécessitent plusieurs étapes de raisonnement pour y répondre (Baillifard et Thurre-Millius, 2020).

2. Une mise en place brutale des évaluations à distance

La crise sanitaire de 2020 est à l'origine d'une généralisation brutale de l'enseignement et de l'évaluation à distance dans les formations universitaires françaises. Ainsi, de mars à fin août 2020, les étudiants de Pharmacie de l'Université de Limoges ont terminé leur parcours de formation à distance et réalisé leurs examens de fin d'année en ligne.

Les enseignants devaient assurer la continuité pédagogique en ligne en employant la plateforme d'enseignement à distance et les outils numériques mis à leur disposition. Pour cela, le département DUENES, composé d'enseignants, d'ingénieurs pédagogiques et de techniciens multimédia les a accompagnés dans la mise en place de leurs cours et des évaluations à distance. Pour chaque parcours de formation, un espace examen a été mis en place par la cellule interdisciplinaire. Les épreuves « Travaux à rendre », « Tests non rédactionnels » et « Évaluations orales » étaient paramétrées en

amont par l'équipe afin que les enseignants se concentrent uniquement sur la création de leurs questions.

Certaines formations (1^{er}, 2^{ème} et 3^{ème} cycles) de Pharmacie suivent un système de notation spécifique. Leurs épreuves de type « tests non rédactionnels » en ligne devaient se rapprocher du système de notation en présentiel. Pour cela, les enseignants avaient pour consigne d'employer les questions de type « Cloze » sur Moodle et d'exclure les questions de type « QCM » et « vrai-faux » dans leurs évaluations. Chaque question devait comporter cinq items. Chaque item contenait trois propositions (vrai, faux, je ne sais pas), une notation spécifique (+ 0.2 point pour une réponse juste à un item, - 0.1 point pour une réponse fautive à un item, 0 point pour le choix « je ne sais pas »). L'ensemble de la question était évalué sur un point. Cependant, la réalisation d'une question « Cloze » est complexe. L'enseignant doit intégrer des lignes de syntaxes (lignes de codes) permettant à la plateforme de déterminer le type de question à générer (QCM, réponse courte, réponse numérique, etc.), les réponses, le nombre de points pour les bonnes et mauvaises réponses ainsi que le feedback. Une formation et une aide technique ont été proposées aux enseignants de Pharmacie pour la création de leurs questions « Cloze ». Un modèle de question Cloze, paramétré et spécifique pour les épreuves de Pharmacie, était à leur disposition pour créer leurs questions, simplifier la tâche et gagner du temps.

Dans ce contexte particulier, il nous semble important d'analyser le niveau de maîtrise des enseignants dans la création de tests numériques pour comprendre s'il peut être un frein dans le choix du format et du niveau de complexité des questions. Pour cela, notre travail s'appuie sur une enquête réalisée en ligne entre le 23 octobre et le 12 novembre 2020 auprès des enseignants de la Faculté de Pharmacie de l'Université de Limoges. Cette enquête avait pour objectifs de mieux connaître :

- Le profil et les connaissances des enseignants de pharmacie en matière d'enseignement à distance ;
- Leurs usages en tests non rédactionnels avant la crise sanitaire du COVID-19, pendant le confinement et lors de la mise en place des examens de fin d'années 2019-2020 à distance.

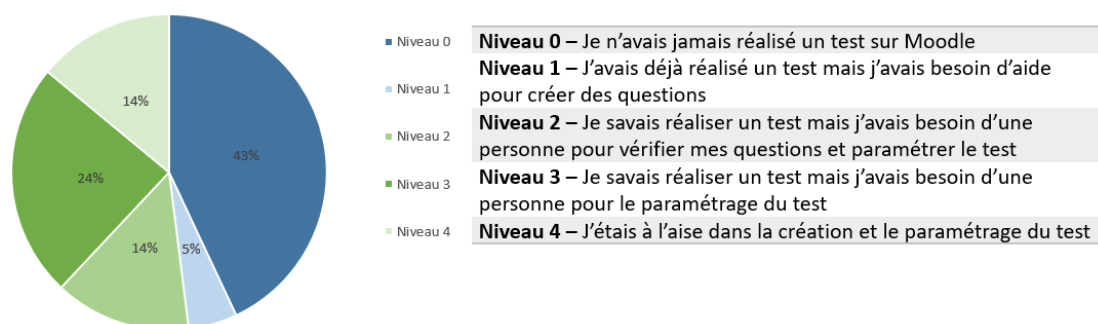
L'enquête a été adressée aux 50 enseignants de Pharmacie (tous statuts confondus) ayant donné des cours dans les parcours de formation de Pharmacie et/ou ayant participé aux examens de fin d'étude pendant l'année 2019-2020. L'étude s'appuie sur les données des 21 enseignants de Pharmacie ayant participé à l'étude.

3. Résultats

3.1. Profil des enseignants

Les résultats de l'enquête mettent en évidence une hétérogénéité du niveau de connaissances des enseignants en matière de tests non rédactionnels sur Moodle. Avant le confinement lié à la COVID-19, 43 % des participants à l'étude n'avaient jamais mis en place de test sur Moodle (niveau 0). Les résultats révèlent que 5 % avaient un niveau 1 concernant Moodle, 14 % un niveau 2, 24 % un niveau 3 et 14 % un niveau 4.

Figure 2 - Niveau des enseignants sur Moodle avant le confinement

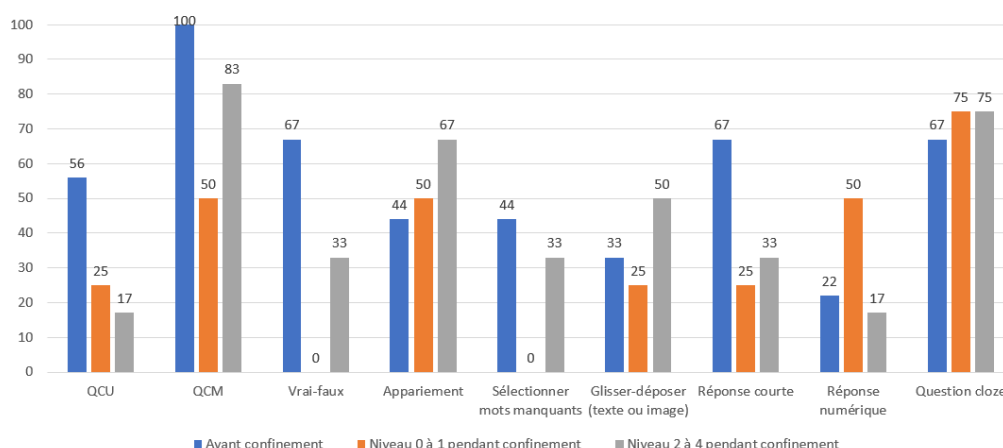


Nous notons également une proportion équitable entre les enseignants ayant un faible niveau (niveaux 0 et 1) dans la création de tests non rédactionnels sur Moodle et les enseignants ayant un niveau avancé (niveau 2 à 4).

3.2. Formats employés dans les tests

Avant le confinement de mars 2020, 43 % des enseignants ayant participé à l'étude utilisaient la plateforme Moodle pour réaliser des tests formatifs. Aucun enseignant de niveau 0 à 1, n'avaient réalisé de tests formatifs sur Moodle ou sur d'autres outils (Socrative, Wooclap, SIDES, ou autre). Tous les enseignants de niveau 2 à 4 employaient avant le confinement, les questions de type QCM. Ils utilisaient également les questions de type vrai-faux (67 %), les questions à réponse courte (67 %), les questions cloze (67 %) et les QCU (56 %).

Figure 3 - Evaluations formatives avant et pendant le confinement



Pendant le confinement (mars à mi-mai 2020), les étudiants terminaient leurs cours à distance et entraient en phase de révision pour leurs examens de fin d'années (de mai à début juillet). Lors de cette période, les enseignants de niveau 0 à 1 utilisaient principalement les questions cloze (75 %), les QCM (50 %), les questions appariement (50 %) et numériques (50 %). Les formats vrai-faux et sélectionner les mots manquants n'ont pas été utilisés.

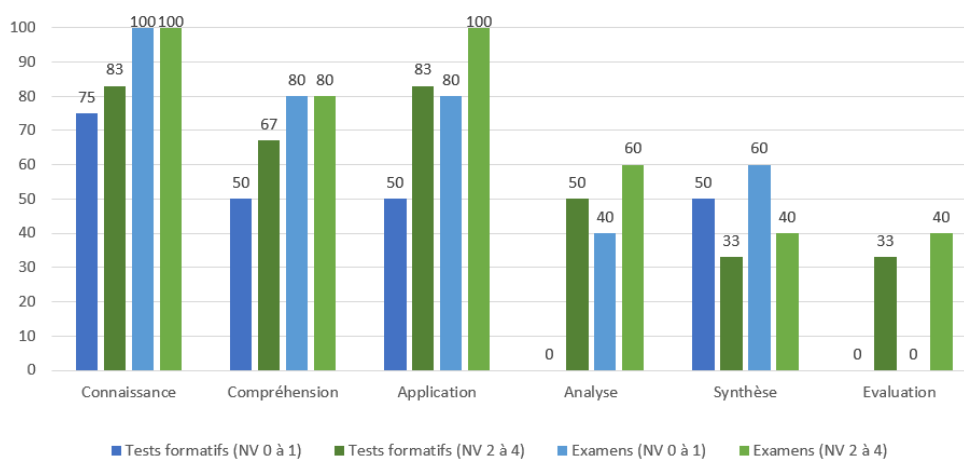
Les enseignants de niveau 2 à 4 proposaient une plus grande variété de questions à leurs étudiants allant des plus simples aux plus complexes (Glisser-déposer, cloze, numérique, etc.). Les questions de type

QCM (83 %), d'appariement (67 %), cloze (67 %) et les glisser-déposer sur un texte ou sur une image (50 %) sont les plus utilisés.

3.3. Niveaux de complexité

Pendant le confinement, les enseignants ayant un niveau 0 à 2 dans la création de tests sur Moodle employaient principalement des questions de niveau de complexité connaissance (75 %), puis dans une moindre mesure, des questions de niveaux compréhension (50 %), application (50 %) et synthèse (50 %). Aucun enseignant de ce groupe n'employait de question de niveaux analyse et évaluation. Les enseignants de niveau 2 à 4 proposaient des questions de tous niveaux de complexité. Ces derniers employaient principalement des questions de niveaux connaissance (83 %), application (83 %), compréhension (67 %) et analyse (50 %).

Figure 5 - Niveaux de complexité dans les tests formatifs pendant le confinement



Les résultats concernant les examens de fin d'année montrent une légère différence par rapport à ceux des évaluations formatives pendant le confinement. Les enseignants de niveau 0 à 1 employaient principalement des questions de niveaux connaissance (100 %), compréhension (80 %), application (80 %) et synthèse (60 %) comme lors des tests formatifs pendant le confinement. Aucun enseignant de ce niveau réalisait de questions de niveau évaluation. Les enseignants de niveau 2 à 4 employaient des questions de tous niveaux de complexité. Nous remarquons une légère augmentation de leurs usages dans les différents niveaux de complexité.

4. Discussion

L'objectif de cette recherche était d'étudier le niveau de maîtrise des enseignants dans la création de tests non rédactionnels sur Moodle et découvrir l'impact de celui-ci dans le choix du format et du niveau de complexité des questions. Notre recherche exploratoire s'est donc portée sur un échantillon réduit d'enseignants de l'Université de Limoges. De ce fait, nous ne prétendons pas à la représentativité de notre échantillon à l'échelle nationale mais présentons la stratégie pédagogique mise en place par notre faculté pour assurer l'évaluation des connaissances à distance et son impact sur le choix des formats de questions employés par les enseignants de Pharmacie lors des différentes évaluations à distance.

L'étude montre que les formats de questions traditionnels (QCM, QCU, vrai-faux, réponse courte) restent les plus utilisés dans les évaluations formatives sur Moodle. Les résultats révèlent également une forte utilisabilité du format Cloze permettant d'intégrer sur la même page différents formats traditionnels (QCM, QCU, appariement et réponse courte) et de se rapprocher au plus près du système de notation des épreuves en présentiel. En appliquant le modèle de question Cloze et son système de notation spécifique, la Faculté de Pharmacie veut également tenir compte de l'effet du hasard dans la note. La littérature scientifique montre une réflexion internationale de la notation dans les évaluations en ligne (Burton, 2001 ; Lesage, Valcke et Sabbe, 2013, etc.) et ses effets sur le facteur chance. Lesage et ses collaborateurs (2013) révèlent que de nombreux tests de type QCM ont été conçus en utilisant la méthode du « Number right scoring ». Les réponses correctes sont notées positivement mais les réponses incorrectes ou absentes ont une valeur à zéro. Ainsi, la somme des notes positives aux questions correspond à la note de l'évaluation. Cependant, en cochant au hasard ou en devinant, les étudiants peuvent répondre positivement. Pour corriger les suppositions et l'effet de hasard, Kurz (1999) propose le modèle « rights minus wrongs » qui pénalise l'étudiant pour toute réponse incorrecte. L'objectif de ce système de notation est de faire comprendre aux apprenants qu'ils perdront des points en répondant incorrectement. Ceci permettrait d'augmenter la fiabilité et la validité du test tout en reflétant la capacité de l'apprenant (Kurz, 1999). D'autres chercheurs soulignent le fait que le marquage négatif ne résout pas le problème de hasard ou de devinette mais introduit un problème de stratégie comportementale des apprenants, non observé dans le modèle du « Number right scoring ». Ainsi, le modèle de question Cloze de la Faculté de Pharmacie de Limoges permet de se rapprocher du système de notation appliqué en présentiel mais également de réduire l'effet de hasard et les stratégies comportementales lors des évaluations.

Par ailleurs, nous souhaitons découvrir l'impact du niveau de maîtrise des enseignants dans la création de tests sur Moodle sur le choix des formats des questions et des niveaux de complexité dans les évaluations. Si notre étude ne nous permet pas de valider ou d'infirmer nos deux dernières hypothèses en raison de la taille réduite de notre échantillon, elle nous donne cependant des informations intéressantes sur les facteurs impliqués dans le choix du format des questions. Nous pouvons en effet remarquer une évolution des usages des enseignants de niveau de maîtrise 2 à 4 dans la création de tests sur Moodle pendant les périodes formatives (avant et pendant le confinement). Les résultats mettent en évidence une forte diminution de l'emploi de questions QCU, vrai-faux et réponse courte pendant la période de confinement (fin des cours et approche des examens). Nous en déduisons que ce changement d'usage est lié à la préparation des examens de fin d'année et à la stratégie d'évaluation sélectionnée par l'équipe pédagogique de la Faculté de Pharmacie de Limoges. Ainsi, il nous apparaît que le choix des formats de questions employés dans les tests non rédactionnels est lié à plusieurs facteurs : (1) l'objectif de l'évaluation (diagnostic, formatif, sommatif), (2) l'objectif pédagogique (niveau de complexité et de difficulté), (3) le niveau technique de l'enseignant sur Moodle, (4) les contraintes pédagogiques (public, lieu de passation, système de notation, moyens techniques à disposition, stratégie pédagogique, etc.). Cependant, le niveau de connaissances techniques des enseignants en matière de numérique peut être compensé par une aide humaine et un support pédagogique. Les formations au numérique et la mise à disposition d'un modèle de question Cloze permettent à tout enseignant de créer plus facilement leurs

épreuves et d'employer un format de question complexe demandant l'intégration de syntaxes pour générer les questions et la notation.

Conclusion

En conclusion, les formats de questions plus traditionnels restent les plus utilisés dans les évaluations formatives à distance. Lors des évaluations sommatives, les enseignants de Pharmacie de l'Université de Limoges emploient principalement le format de question Cloze, plus complexe dans sa mise en œuvre, afin d'obtenir un système de notation uniforme et connu des étudiants. Cependant, il nous semble important de former davantage les enseignants aux outils et à la création de ressources et d'activités numériques.

Références

- Bloom, B. S. (1956). *Taxonomy of educational objectives: The classification of educational goals*. New York: Longmans, Green.
- Burton, R. F. (2001). Quantifying the Effects of Chance in Multiple Choice and True/False Tests: Question Selection and Guessing of Answers. *Assessment & Evaluation in Higher Education*, 26(1), 41–50.
- Cartwright, G.F. (1975). A promising innovation: computer assisted test construction. *Learning and Development*, 6, (4), 1-3.
- Cartwright, G.F., Derevensky, J. (1975). An attitudinal Study of Computer Assisted Testing as a learning Method.
- Dudley, T.J. (1973). How the computer assists in pacing and testing students progress. *Educational Technology*, 13, 21-23.
- Kelly, F. J. (1916). The Kansas Silent Reading Tests. *Journal of Educational Psychology*, 7(2), 63–80.
- Kurz, T.B. (1999). *A review of scoring algorithms for multiple-choice tests*. Paper Presented at the Annual Meeting of the Southwest Educational Research Association, San Antonio, TX
- Lesage, E., Valcke, M. and Sabbe, E. (2013). Scoring Methods for Multiple Choice Assessment in Higher Education: Is it Still a Matter of Number Right Scoring or Negative Marking? *Studies in Educational Evaluation*, 39(3), 188–193.
- Sousa, D.A. (2002). *Un cerveau pour apprendre : Comment rendre le processus enseignement-apprentissage plus efficace*. Montréal : Chenelière Education.