



## Une analyse automatisée des modalités d'évaluation des MOOC de la plate-forme FUN

Matthieu Cisel

**Résumé :** Nous nous penchons sur les modalités d'évaluation d'une centaine de MOOC de la plate-forme France Université Numérique pour chercher à appréhender dans quelle mesure celles-ci pourraient contribuer à expliquer les faibles taux de certification qui y sont observés. Sur la base de plusieurs milliers d'exercices, nous avons recueilli de manière automatisée des données sur le type d'évaluation mobilisée – évaluations par les pairs, évaluations automatisées – ainsi que sur leur nombre ou sur leur paramétrage, comme le nombre d'essais autorisés par exercice. Nous suggérons que si la charge de travail peut varier considérablement d'un MOOC à l'autre, les exercices se comptant parfois par centaines, le niveau d'exigence de ces évaluations est généralement faible, car elles ne sont pas conçues pour mettre en difficulté les apprenants. Dans ces conditions, il est peu vraisemblable que les faibles taux de certification observés puissent s'expliquer par une incapacité d'une part significative des apprenants à réussir avec succès ces évaluations.

**Mots-clés :** MOOC, évaluation, ingénierie pédagogique.

**Abstract:** We focus on how assessment was performed in more than a hundred MOOCs of the French platform France Université Numérique in order to better comprehend to which extent MOOCs' low completion rates could be a consequence of assessment policies. We gathered automatically detailed data on which type of assessment was favored – peer assessment, automatized assessment – as well as their number and their settings, like the number of attempts allowed by exercise. We suggest that, if the workload is sometimes heavy, with hundreds of exercises sometimes required to obtain the certificate, the level of requirement is low, since evaluation is apparently not meant to challenge learners in these MOOCs. Therefore, it is unlikely that their low completion rates could be explained by the inability of a significant proportion of learners to succeed in passing assessment tests.

**Keywords:** MOOC, evaluation, instructional design.



## Introduction

Alors que le débat sur les faibles taux de certification des MOOC bat son plein, la fondatrice de la plate-forme américaine *Coursera* publie sur la question un réquisitoire qui fera date (Koller *et al.*, 2013), dont ces quelques phrases d'introduction, gardées dans leur langue d'origine, donnent le ton : « To a university professor accustomed to the traditional audience of committed, paying students in a brick-and-mortar classroom, the image of continuously-emptying lecture halls — where only one in every 20 students remains to the end — is an understandably frightening prospect. But is this really the appropriate framework for thinking about student success in MOOCs? »

Les auteurs présentent un ensemble d'hypothèses susceptibles d'expliquer la faiblesse des taux de certification observés : inscription à de multiples cours, diversité des usages, existence de cours superposés. Curieusement, une hypothèse des plus simples manque à l'appel. En effet, un certain nombre de participants sont susceptibles d'échouer à obtenir le certificat car ils ne parviennent pas à réaliser avec succès les activités évaluées. Cette hypothèse, que nous qualifierons d'hypothèse de l'échec académique en référence à la terminologie utilisée dans les travaux séminaux sur la persistance dans les universités américaines (Tinto, 1975), est d'autant plus plausible que les cours de la plate-forme *Coursera* sont significativement plus exigeants en termes de charge de travail que la moyenne des MOOC (Cisel, 2016). Nous définirons de manière opératoire l'échec académique comme le fait de ne pas parvenir à obtenir une note suffisante pour obtenir le certificat, et ce bien que les activités évaluées requises pour son obtention aient été réalisées, toutes les éventuelles opportunités de refaire un exercice raté ayant été saisies. Plusieurs considérations légitiment à nos yeux la réalisation d'analyses des modalités d'évaluation par la communauté scientifique.

Tout d'abord, des enquêtes montrent qu'une partie conséquente des apprenants mettent sur leur Curriculum Vitae les certificats des MOOC obtenus (Cisel, 2016), ce qui ne manque pas de susciter un certain nombre d'interrogations quant à ce que ces certificats représentent en termes de validation d'acquis. La réflexion sur le sujet est sévèrement amputée par la rareté des analyses empiriques portant sur le travail nécessaire à l'obtention de ces certificats.

Se pose ensuite la question de la persistance au sein de ces formations, qui n'a pas manqué d'attirer l'attention de la communauté scientifique. D'une part les modèles théoriques comme le modèle de rétention de Tinto (1975) et ses adaptations successives au contexte de la formation à distance suggèrent que les performances en termes de notes, a priori corrélées avec la difficulté des évaluations, ont un impact important sur la persistance. D'autre part, des travaux empiriques consacrés aux MOOC (Jordan, 2014, 2015) suggèrent que leurs taux de certification dépendent des

modalités d'évaluation mises en place, les cours mobilisant l'évaluation par les pairs affichant notamment des taux de certification significativement plus bas que la moyenne.

À notre connaissance, peu de recherches se sont penchées spécifiquement sur la question des modalités d'évaluation dans les MOOC. Soit l'évaluation est abordée au prisme de la docimologie (Kulkarni et al., 2013), soit les modalités d'évaluation ne constituent dans les quelques études consacrées aux caractéristiques des MOOC (Jordan, 2014, 2015 ; Toven-Lindsay *et al.*, 2015 ; Margaryan *et al.*, 2015) qu'une dimension parmi d'autres, traitée de manière relativement superficielle du fait du caractère manuel des analyses.

Nous cherchons ici à contribuer à combler cette lacune au travers de l'analyse de la plate-forme française France Université Numérique (FUN), pour tâcher d'alimenter la réflexion sur une problématique qui englobe plusieurs sous-questions : dans quelle mesure l'attrition observée dans les MOOC est-elle influencée par les choix réalisés par les équipes pédagogiques en termes de modalités d'évaluation ? S'il est de nombreuses manières de traiter cette problématique, nous ne nous focaliserons ici que sur la seule question des caractéristiques des dispositifs, nous limitant ce faisant aux questions suivantes : Sur la base de quelles activités reposent l'évaluation dans les MOOC ? Dans quelle mesure le paramétrage de ces évaluations favorise-t-il, ou au contraire défavorise-t-il l'échec académique au sein des MOOC ? Dans quelle mesure le niveau d'exigence des évaluations est-il lié à la charge de travail associée au cours ?

Nous faisons dès à présent l'hypothèse selon laquelle le niveau d'exigence de ces MOOC est relativement bas, et que les évaluations sont conçues de sorte à ne pas mettre en difficulté les participants, rendant peu probable l'existence de l'échec académique tel que nous l'avons défini plus tôt. En revanche, le nombre d'évaluations à réaliser peut être conséquent, favorisant alors le retrait volontaire, qui correspond à une décision positive de quitter la formation.

Pour mettre à l'épreuve ces hypothèses, nous avons réalisé de manière automatisée l'analyse de la structure de plus d'une centaine de cours, qui s'apparentent à des xMOOC, c'est-à-dire qu'ils ne correspondent pas aux cours connectivistes majoritaires dans la première génération de MOOC (Daniel, 2012). Notre propos est également de nature méthodologique, dans la mesure où nous cherchons à mettre en évidence l'intérêt de l'approche qui consiste à analyser automatiquement des dispositifs. Si celle-ci a des limites, elle permet non seulement de caractériser un nombre important de cours, mais aussi de manière plus détaillée que ne le permet une analyse manuelle.

Au-delà de la question de l'influence des modalités d'évaluation sur le comportement des apprenants, les éléments récoltés lors de cette analyse peuvent nous renseigner quant à la logique qui sous-tend le dispositif. Il est important de noter à cet égard que l'analyse présentée ici est

concomitante d'une enquête auprès de concepteurs de MOOC visant à mieux appréhender les motivations des enseignants qui s'engagent dans la mise en place de ces formations. Les entretiens semi-directifs menés visaient à corroborer les interprétations que nous avons faites des analyses quantitatives réalisées ici, mais le volet qualitatif de cette approche de méthodes mixtes ne saurait être présenté dans cet article faute de place.

Dans la mesure où les réflexions menées ici sur les modalités d'évaluation ont comme origine les travaux réalisés dans le cadre de l'attrition en formation, et en particulier le modèle de rétention de Tinto (1975), nous nous proposons de revenir brièvement sur l'émergence de ce cadre d'analyse avant de présenter certains des travaux portant sur les MOOC, que nous avons jugé pertinents au regard des questions qui structurent cette contribution.

## **Revue de la littérature**

Des recherches sur l'attrition dans l'enseignement supérieur nord-américain, nous retiendrons une dichotomie particulièrement opératoire pour notre propos, comme le rappellent Henderikx, Kreijns et Kalz (2017) dans leurs travaux sur la typologie des utilisateurs de MOOC, à savoir la distinction entre le retrait volontaire et l'échec académique (Tinto, 1975). Le premier est défini comme une décision positive de l'apprenant de se retirer de la formation. L'échec académique est en revanche une situation subie ; il correspond à l'incapacité de l'étudiant à remplir les critères d'obtention du diplôme, notamment du fait de notes insuffisantes.

Comme le souligne Tinto (2006), cette dichotomie est apparue à une époque où « les étudiants échouaient, et non les institutions ». Ceux qui ne persistaient pas au sein de la formation étaient considérés comme moins capables, moins motivés, et moins enclins à reconnaître les bénéfices que l'on pouvait retirer d'une formation universitaire. Cette approche, qui domine pendant plusieurs décennies, finira par évoluer sous l'impulsion de chercheurs comme Tinto.

Tinto (1975), dans la lignée des travaux de Spady (1970) et Rootman (1972), développe un modèle de rétention pour appréhender les mécanismes du retrait volontaire. Son modèle postule que la rétention, comprise ici comme l'inverse de l'attrition, dépend avant tout des deux paramètres que sont l'engagement vis-à-vis des objectifs (Goal Commitment), et l'engagement vis-à-vis de l'institution (Institutional Commitment).

Cette approche ne nie pas l'importance des caractéristiques des étudiants. Soulignons en effet que le *background* correspond aux motivations pour s'inscrire, au contexte socioprofessionnel, aux expériences antérieures, etc. Néanmoins, elle a l'avantage de prendre en compte les caractéristiques du dispositif. L'intégration sociale correspond ainsi aux liens résultant de l'interaction quotidienne

des étudiants avec leurs pairs ou avec les équipes pédagogiques. Tinto adopte une approche interactionniste dans la mesure où il accorde une importance considérable à l'intégration sociale. Il ne néglige néanmoins pas l'intégration académique, qui dépend notamment des performances, comme les notes obtenues, et donc des modalités d'évaluation des étudiants. C'est cette dernière composante du modèle, et en particulier sur les performances en termes de notes, qui nous a poussé à mettre en œuvre les analyses présentées dans cet article.

Comme nous le soulignons plus tôt, il est plausible que les notes des apprenants soient affectées de manière significative si le niveau d'exigence des évaluations est trop élevé. Cette situation pourrait contribuer à augmenter l'échec académique si les notes obtenues sont éliminatoires. Avant de nous pencher sur les éléments de la littérature MOOC susceptibles d'éclairer ce phénomène, revenons sur la question de l'adaptation du modèle de Tinto au contexte de la formation à distance.

Les forts taux d'attrition constatés au sein des formations à distance vont conduire un certain nombre d'auteurs à appliquer et adapter les modèles développés par Tinto à ce contexte particulier (Bean & Metzner, 1985 ; Thomas, 2000 ; Kemp, 2002 ; Rovai, 2003). Les adaptations requises sont de plusieurs ordres ; tout d'abord, on ne raisonne plus à l'échelle d'une année universitaire, mais à celle d'un cours isolé, n'impliquant souvent aucun contact direct, en face à face, avec l'enseignant. L'influence des caractéristiques du cours devient mécaniquement plus importante que dans le cadre d'un cycle d'études de quatre ans.

Bien que de nouvelles variables aient été prises en compte, les composantes que sont l'intégration académique et l'intégration sociale continuent à constituer des éléments centraux de la plupart de ces modèles. Nous nous intéresserons à travers la question de l'évaluation à une seule des composantes du modèle de rétention de Tinto (1975), les performances mesurées en termes de notes. Ce faisant, nous traiterons des questions de persistance, définie ici comme la proportion des actions prescrites (visionnage de vidéos, réalisation d'activités évaluées, etc.) effectivement réalisées par l'apprenant.

Il est intéressant de constater, au fil des revues sur les facteurs expliquant l'attrition dans la formation à distance (Bourdages et Delmotte, 2007 ; Lee & Choi, 2011), que la question des modalités d'évaluation semble peu étudiée. Ces revues de littérature reprennent systématiquement la dichotomie entre impact des caractéristiques des apprenants, et variables institutionnelles. Les premières regroupent les variables démographiques, environnementales (Brown, 1996) - le contexte dans lequel s'inscrit l'apprenant – mais aussi des variables du processus d'apprentissage, comme le style d'apprentissage ou la capacité à autoréguler ses apprentissages (Joo et al., 2012). Les secondes regroupent l'interaction entre équipe pédagogique et étudiants – l'intégration sociale de Tinto (1975) – et en particulier l'aide qui peut être apportée sous forme de tutorat. D'après la revue réalisée par

Bourdages et Delmotte (2007), les travaux portent le plus souvent sur l'efficacité de tel ou tel type d'intervention sur la persistance des étudiants, plus que sur l'impact des caractéristiques des dispositifs en matière d'évaluation.

Cette dernière question est davantage traitée dans les travaux réalisés en psychologie sur l'impact positif des exercices, et notamment des exercices automatisés (Roediger et Karpicke, 2006 ; Zaromb et Roediger, 2010), sur la mémorisation, et sur le processus d'apprentissage en général. Agarwal *et al.* (2008) analysent en particulier l'efficacité de l'approche « à livre ouvert », où l'on peut réaliser l'exercice tout en ayant accès aux ressources sur lequel se fonde l'évaluation, situation qui correspond à la plupart des MOOC. Si la question des exercices automatisés est traitée sous l'angle de l'effet test, ou *testing effect* pour les anglo-saxons, sur la mémorisation du contenu du cours, la question de la persistance au sein de la formation semble largement sous-investie.

Les travaux consacrés à la persistance dans les MOOC se penchent avant tout sur les caractéristiques des apprenants, comme la motivation pour s'inscrire (Greene, Oswald et Pomerantz, 2015) et les capacités à s'autoréguler (Kizilcec *et al.*, 2017). Néanmoins, notre objectif n'est pas ici de recenser les nombreuses enquêtes consacrées aux caractéristiques des apprenants de MOOC. Ce choix réside dans le fait que nous nous proposons de transférer la focale sur les caractéristiques des dispositifs, que peu d'auteurs ont étudié et croisé avec la question du comportement des apprenants.

Seules quelques exceptions notables sont dignes d'être relevées. Ainsi, Ma *et al.* (2015) et Alraimi *et al.* (2015) pour les MOOC, font le lien entre persistance et interventions de l'équipe pédagogique sur les forums de discussion, ou réputation de l'institution organisatrice. Là encore, les auteurs, s'ils font preuve d'originalité de par leur objet d'étude, se penchent sur des variables déjà traitées dans la littérature sur la formation à distance, et font fi de la question des modalités d'évaluation. À défaut de pouvoir croiser dans cet article caractéristiques des dispositifs et comportement des utilisateurs, nous avons cherché à caractériser de manière aussi détaillée que possible la structure des cours ainsi que les modalités d'évaluation, selon une méthodologie qu'il nous faut maintenant présenter.

## **Méthodologie**

Dans les paragraphes qui suivent, nous allons présenter les différentes informations que nous avons extraites à partir de la structure d'un dispositif. Ce travail a été précédé d'une enquête qualitative auprès de trente concepteurs de MOOC, enquête dont nous ne présenterons ici ni les résultats ni la méthodologie, concision oblige, mais qui a contribué à l'interprétation des résultats. Nous allons notamment nous intéresser à ce que nous nommerons le niveau d'exigence de la formation, dont nous avons défini deux dimensions sur lesquelles nous allons revenir : la charge de travail d'une part, et le niveau d'exigence des évaluations d'autre part.

Par structure, nous entendons ici les caractéristiques du dispositif observables à partir d'un export du MOOC depuis la plate-forme. Depuis toute plate-forme fondée sur la technologie *Open edX*, ce qui est le cas de *FUN*, il est possible de télécharger une version du cours pour archivage. Cette archive compressée contient un certain nombre d'éléments relatifs au paramétrage du cours – nombre et identification des composantes de la formation, paramétrages de l'évaluation, etc. – éléments qui composent ce que nous avons appelé la structure du cours, et que nous avons obtenue selon la méthode qui suit.

**Tableau 1 : Un récapitulatif des données récupérées dans 131 MOOC par le web scraper**

Données récoltées sur la structure des cours présentées dans l'article
Nombre d'évaluations présentes dans le dispositif
Nombre d'évaluations par les pairs
Nombre de <i>Questionnaires à Choix Multiples</i>
Nombre de <i>Questionnaires à Choix Unique</i>
Nombre d'exercices de type <i>Réponse Numérique</i>
Nombre d'exercices de type <i>Texte à trou</i>
Nombre d'essais autorisés par exercice (pour les exercices automatisés)
Seuil de passage, proportion de la note maximale permettant l'obtention du certificat
Nombre de vidéos présentes dans le dispositif

Les structures des cours ont été analysées automatiquement par des programmes *ad hoc*, récoltant les données relevées dans le Tableau 1. Les données relatives à la structure des dispositifs ont été récoltées au format JSON via un accès à une API dédiée, après que des droits d'accès nous aient été accordés par les administrateurs de la plate-forme. À chaque cours correspond un fichier de données JSON dont les informations suivantes sont extraites via un programme *ad hoc*: nombre de chapitres, nombre de vidéos, nombre et type d'exercices automatisés, nombre d'activités évaluées par les pairs, dates de lancement des différents chapitres, nombre d'essais autorisés par exercice automatisé, seuil de validation ou seuil de passage. Seuls les MOOC dont la structure était totalement stabilisée ont été analysés, ce qui revient à ne prendre en considération que les seuls cours dont au moins une itération a été terminée. En prenant garde de n'analyser qu'une seule itération par cours, nous nous sommes intéressés à 131 MOOC distincts, organisés entre janvier 2014 et octobre 2015.

Comme Jordan (2015) nous avons cherché à qualifier les modalités d'évaluation des apprenants, en distinguant d'une part les évaluations automatisées, et d'autre part les évaluations réalisées par les

pairs. Les évaluations automatisées correspondent aux tests corrigés de manière automatique, sans rétroaction de la part d'un humain sur la réponse fournie par l'apprenant. Il en existe une certaine diversité au sein de la plate-forme FUN<sup>u</sup>, fondée sur la plate-forme *Open edX*. Nous en retiendrons quatre formes : questions à choix multiples (QCM), questions à choix unique (QCU), réponses numériques, et textes à trous. Dans le cas des réponses numériques, l'apprenant doit fournir une valeur numérique correspondant à une valeur précise, ou, à défaut, comprise dans une fourchette définie par le concepteur du cours. Nous considérerons que plus un MOOC contient d'activités évaluées, plus la charge de travail correspondant au certificat est élevée, toutes choses restant égales par ailleurs.

Nous nous intéressons ensuite à la proportion d'exercices de textes à trou ou de type *réponse numérique*, et inversement, de la proportion de QCU et de QCM au sein des évaluations automatisées. Les QCU ou des QCM correspondent généralement à des questions factuelles sur le contenu du cours, là où la réponse numérique correspond généralement à une application du cours et implique vraisemblablement un investissement plus important. Néanmoins, une étude plus approfondie, qui sortirait du cadre de notre propos, serait nécessaire pour étayer cette hypothèse, dont la démonstration n'est pas nécessaire à la suite de notre propos.

Nous définirons ce que nous nommerons le niveau d'exigence des évaluations sur la base du paramétrage de ces mêmes évaluations. Le premier paramètre relatif au niveau d'exigence des évaluations est le nombre d'essais autorisés pour une évaluation automatisée. Lorsqu'un apprenant donne une mauvaise réponse à une telle évaluation, il peut se voir offrir la possibilité de recommencer l'exercice en question, ce qui n'est a priori pas le cas pour l'évaluation par les pairs. Par essence, plus le nombre d'essais est élevé, plus il est facile de trouver la réponse à un exercice, en particulier pour les QCU et QCM. Par conséquent, nous considérerons que le niveau d'exigence des évaluations est d'autant plus faible que le nombre d'essais autorisés est élevé. Cette considération nous amènera à nous pencher sur deux métriques, d'une part le nombre moyen d'essais autorisés par exercice, et d'autre part la proportion des exercices automatisés n'autorisant qu'un seul essai. Le niveau d'exigence des évaluations est positivement corrélé à la première métrique, et négativement à la seconde.

Enfin, le second indicateur qui a attiré notre attention est le seuil de validation, qui correspond à la proportion de la note maximale en-deçà de laquelle il n'est pas possible d'obtenir l'attestation de réussite avec succès. Nous parlerons de certificat pour désigner cette attestation, par souci de concision. Rappelons qu'il est gratuit, et que sur la période que couvrent ces analyses, il n'y avait pas de vérification de l'identité de l'apprenant. Ce seuil est donc compris entre 0 et 1 ; nous considérerons que plus il se rapproche de 1, plus le niveau d'exigence des évaluations est élevé.

Notons que notre manière de procéder ne permet pas d'appréhender le MOOC dans son ensemble, mais uniquement la partie du dispositif présente sur *FUN*. Il arrive que certains MOOC mobilisent plusieurs plates-formes pour compenser l'absence d'une fonctionnalité critique sur une plate-forme donnée. Ainsi, Khaneboubi et Baron (2015) présentent une analyse de données collectées notamment à partir d'une plate-forme *Moodle* utilisée dans le cadre d'un MOOC organisé sur *FUN*. Les activités qui s'y déroulent, évaluées ou non, ne peuvent être prises en considération dans l'analyse d'une structure de MOOC telle que nous l'avons réalisée. Nous négligerons néanmoins cette pratique ; tout d'abord, l'inspection manuelle de la plupart des dispositifs suggère qu'elle est relativement marginale. Ensuite, il est difficile pour le concepteur de rapporter les éventuelles évaluations réalisées par ces moyens au sein de la plate-forme *FUN*, autant pour des raisons techniques que légales, et les éventuelles notes qui ne seraient pas rapportées ne sont pas prises en compte dans l'obtention du certificat ou de l'attestation de réussite. Ce faisant, elles n'ont que peu d'incidence sur les paramètres qui nous intéressent, à savoir l'obtention de cette attestation.

Notons également que dans les analyses descriptives qui suivent, nous négligerons délibérément plusieurs paramètres. Il est par exemple possible que pour un certain nombre de cours, le nombre de vidéos proposées soit inversement proportionnel à la durée moyenne de ces vidéos, de sorte qu'un nombre de vidéos plus élevé n'implique pas nécessairement une durée totale de vidéo plus élevée. Par ailleurs, il est également possible que la consultation de ces vidéos, ainsi que des autres ressources du cours, non prises en compte dans cette étude, soit optionnelle pour la réalisation des exercices, ou qu'ils ne correspondent pas uniquement à des vidéos de cours. Enfin, il est plausible que certains des exercices du dispositif ne soient pas pris en compte pour l'obtention du certificat, et soient également optionnels. Néanmoins, une analyse aussi fine de ces paramètres dépasse largement le cadre de ce travail, et nous considérerons comme négligeables pour notre analyse les biais que ces éléments sont susceptibles d'introduire.

Nous proposerons enfin pour caractériser les dispositifs un certain nombre de statistiques descriptives classiques, comme la moyenne ou l'écart-type, systématiquement introduit par le symbole  $\pm$ . Le modèle linéaire généralisé (dont l'abréviation est GLM), est utilisé à une reprise, car la variable d'intérêt suit une loi de Poisson. Le caractère significatif des tests est mesuré avec la p-value. Ces précisions méthodologiques étant fournies, abordons la présentation des résultats proprement dits.

## **Résultats**

Nous allons successivement nous intéresser à la place réduite qu'occupe l'évaluation par les pairs, puis à la nature des exercices automatisés utilisés, au nombre de tentatives autorisées, et finalement,

aux seuils de validation du cours. Nous concluons sur le nombre de vidéos et d'exercices automatisés compris dans le cours. Nous verrons comment ces différents paramètres sont susceptibles de nous éclairer quant à un éventuel échec académique des utilisateurs des cours.

Notre premier constat est que la plupart des MOOC n'impose pas aux apprenants de soumettre une production qui serait évaluée par les pairs. Seuls 25% des dispositifs analysés imposent de rendre un ou plusieurs devoirs évalués par les pairs, qui peuvent être décomposés en 10% avec seulement une production évaluée par les pairs, 8% avec deux productions, et 7% avec trois productions et plus. Ces MOOC ne correspondent donc pas à des formations à distance qui exigeraient systématiquement une production de la part des apprenants pour valider des acquis. Nous interprétons ce résultat comme le fait qu'ils représentent des dispositifs allégés, adaptés à un public que les concepteurs jugent peut-être réticent à s'engager dans la réalisation de productions nécessitant un investissement temporel trop conséquent. La forte représentation des QCM et des QCU parmi les exercices automatisés nous fait abonder dans ce sens. Nous calculons dans un premier temps leur proportion pour chaque cours, puis nous nous intéressons à la distribution de ces proportions à l'échelle des 131 cours.

Les chiffres qui suivent correspondent à la fréquence moyenne de tel ou tel type d'exercice automatisé à l'échelle de cet ensemble de cours, et le chiffre entre parenthèse à l'écart-type correspondant à cette distribution. Il ressort de cette analyse menée sur 16730 exercices présents dans 131 cours que 56,9% ( $\pm 30,6$ ) des exercices automatisés sont des QCM, 30,6% ( $\pm 32,2$ ) sont des QCU, 2,8% ( $\pm 6,5$ ) des textes à trou, et seulement 1,9% ( $\pm 5,4$ ) des réponses numériques. Le fait que les écarts-types soient de l'ordre de grandeur des fréquences moyennes reflète le caractère hétérogène des dispositifs, certains comportant essentiellement des QCM, d'autres essentiellement des QCU. Les concepteurs de MOOC semblent donc favoriser des exercices avec un nombre fini de réponses possibles, les cours se différenciant essentiellement par la proportion relative de QCM et de QCU. On constate que 84% des MOOC analysés n'exigent aucune réponse numérique de la part des apprenants, et 85% aucun texte à trou ; 69% ne proposent ni l'un ni l'autre.

Poursuivons l'analyse en nous intéressant maintenant au paramétrage fin de ces exercices automatisés, paramétrage qui touche au niveau d'exigence de ces évaluations. Nous constatons que 39% des cours permettent deux tentatives ou davantage par exercice automatisé en moyenne, tandis que dans 70% des dispositifs, moins de 20% de ces exercices n'autorisent qu'une seule tentative (Figure 1A). Les exercices automatisés peuvent le plus souvent être refaits en cas d'erreur, ce qui constitue un indice supplémentaire corroborant notre hypothèse selon laquelle le niveau d'exigence des évaluations des MOOC analysés ici est relativement faible.

Présentons maintenant un dernier élément qui va dans le sens de notre propos, le seuil de validation des cours, relativement bas dans la plupart des formations. La majorité des cours ont un seuil de passage inférieur à 60% de la note maximale (Figure 1B). Dans 50 % des MOOC de FUN, le seuil de passage correspond à seulement 50% de la note maximale ; ce n'est que dans 4 % des cas que le seuil de passage est strictement supérieur à 70%. La conjonction de ces différents résultats - faible proportion des exercices automatisés qui n'autorisent qu'une tentative, seuils de validation relativement bas - semble aller dans le sens de notre hypothèse relative au faible niveau d'exigence de l'évaluation au sein de ces formations.

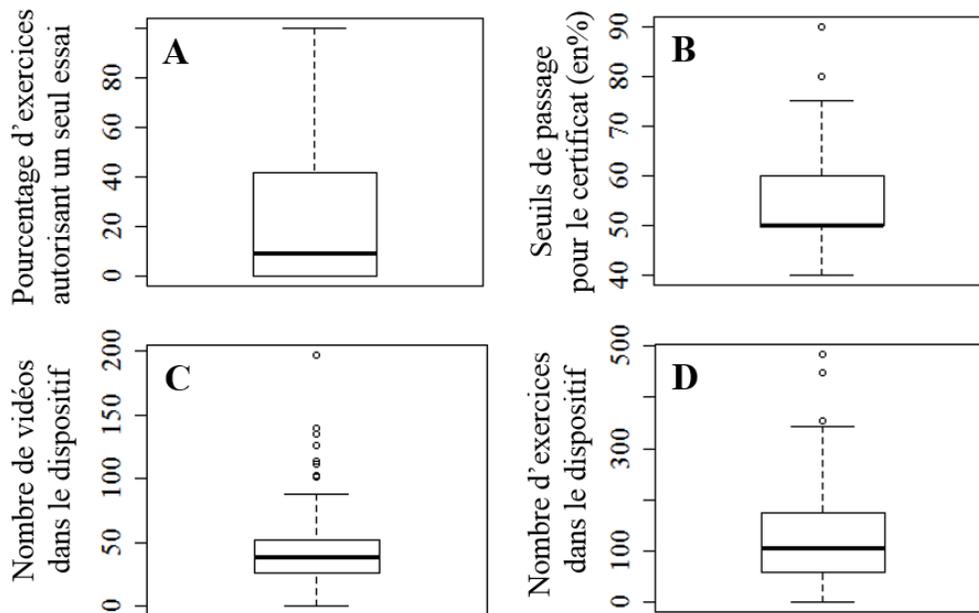


Figure 1 : Distribution de quatre caractéristiques des MOOC pour 131 cours de FUN, relatives au paramétrage des évaluations ou au nombre d'items présent dans le dispositif

Néanmoins, conservons à l'esprit que, toutes choses restant égales par ailleurs, le niveau d'exigence d'un MOOC est fonction de la charge de travail qu'il implique, qui dépend notamment de la quantité d'exercices à réaliser, et de la quantité de ressources à consulter pour pouvoir réaliser ces exercices. Un faible niveau d'exigence moyen pour les exercices automatisés peut être compensé par un nombre élevé d'exercices, et réciproquement. Cette considération nous a amenés à analyser la distribution du nombre d'évaluations comme du nombre de vidéos pédagogiques à l'échelle l'ensemble des dispositifs. L'unité élémentaire est ici l'exercice, indépendamment du type d'exercice dont il s'agit, texte à trou, réponse numérique, QCM ou QCU.

Le nombre de vidéos varie entre moins de quinze à près de deux cents, une hétérogénéité qui met en évidence les différences de charge de travail imposées aux participants (Figure 1C). Si la majorité des dispositifs comportent entre 15 et 50 vidéos, les MOOC comprennent généralement plus d’une cinquantaine d’exercices automatisés (Figure 1D) ; nombreux sont ceux qui en comptent plus d’une centaine, voire plus de quatre cents.

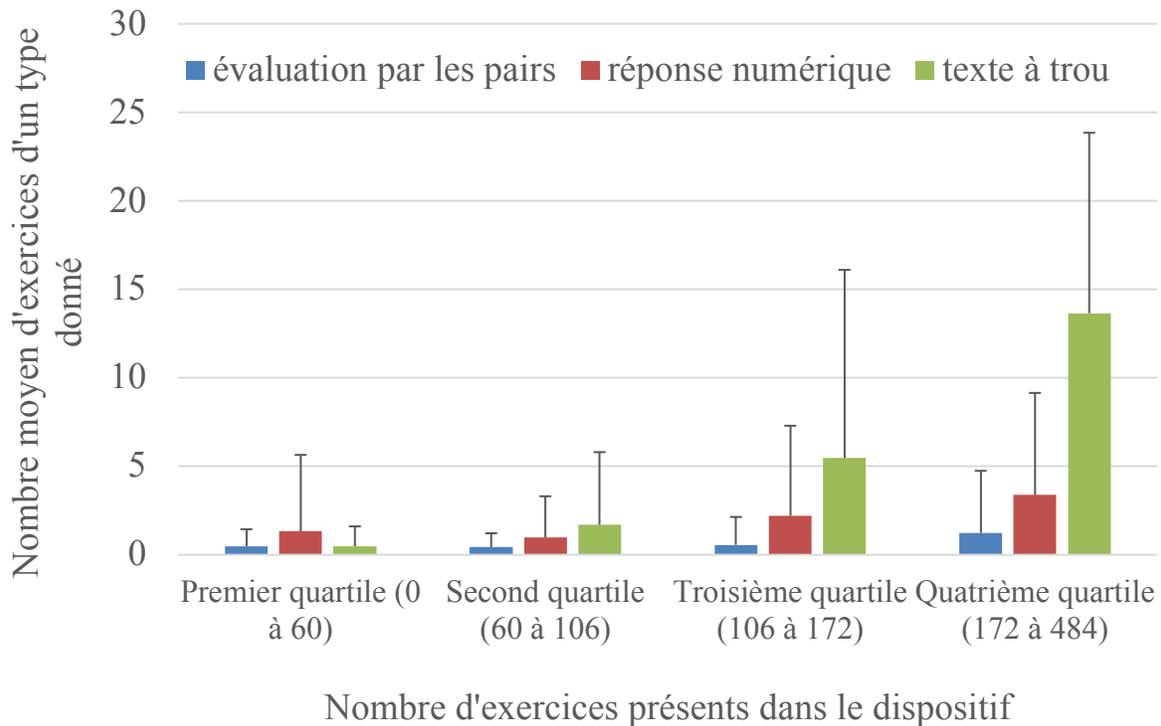


Figure 2 : Évolution de la nature des évaluations en fonction du nombre d'exercices

Nous avons représenté dans la Figure 2 l’évolution du nombre de ces exercices en fonction du nombre total d’exercices, en divisant les dispositifs selon quatre quartiles. On constate que c’est dans le quatrième quartile, lorsque le nombre d’exercices proposés est maximal, que le nombre moyen d’évaluations par les pairs est le plus élevé (GLM, p-value=0,00014). Il en va de même pour les réponses numériques (GLM, p-value<2.10\*), et les textes à trous (GLM, p-value=0,005). Ces résultats suggèrent qu’il existe une forme d’homothétie des structures des dispositifs ; le nombre d’exercices a priori plus exigeants en temps, comme les évaluations par les pairs, augmente en même temps que s’accroît le nombre d’exercices.

## Discussion

Reprenons les unes après les autres les diverses questions relatives à la structure des dispositifs avant de revenir, dans une logique de formulation d'hypothèses, sur le lien entre cette structure et le comportement des utilisateurs. Nous souhaitons dans un premier temps nous pencher sur les activités mobilisées dans l'évaluation. Nous avons constaté que les cours analysés, qui représentent la majorité des MOOC français (Cisel, 2016), sont caractérisés par une évaluation essentiellement fondée sur des exercices automatisés de type QCU ou QCM, et non sur des évaluations par les pairs, ni sur des exercices de type texte à trou. S'il est vraisemblable que ces QCU et QCM portent sur la connaissance factuelle du contenu du cours davantage que sur des exercices d'application, une analyse plus approfondie serait nécessaire pour étayer un tel propos.

Nous nous sommes ensuite penchés sur le paramétrage de ces évaluations, pour montrer que le plus souvent plusieurs tentatives sont laissées aux apprenants, et que les seuils de validation sont relativement bas. Par ailleurs, les vidéos et autres ressources sur lesquelles se fondent ces exercices sont accessibles en permanence, donc y compris pendant la réalisation de ces exercices. La combinaison de ces divers éléments nous amène à avancer l'hypothèse selon laquelle les concepteurs de cours n'ont pas tant pensé l'évaluation pour mettre en difficulté les apprenants, que pour favoriser la mémorisation du contenu, selon une logique relevant de l'effet test (Roediger et Karpicke, 2006) discuté en introduction de cet article.

Si, faute de place, nous ne pouvons présenter ici les recherches qualitatives qui corroborent une telle hypothèse, soulignons qu'un certain nombre d'études quantitatives publiées aussi bien dans la littérature scientifique (Evans & Myrick, 2015) que dans la littérature grise (Kolowich, 2013 ; Hollands & Thirtalli, 2014, 2015) lui donnent du crédit. Ainsi, Kolowich (2013) montre notamment que pour nombre de concepteurs, le certificat délivré ne devrait pas être associé à des crédits universitaires, car les modalités d'évaluation ne sont pas suffisamment exigeantes, un élément cohérent avec nos résultats sur le paramétrage des évaluations.

Pour reprendre les concepts issus des travaux sur la persistance en formation (Tinto, 1975), ce constat relatif au paramétrage corrobore l'hypothèse selon laquelle il est probable que soit marginal le cas de figure de l'échec académique, où des participants échoueraient à obtenir le certificat après avoir épuisé toutes leurs tentatives. En revanche, le retrait volontaire, qui constitue une décision positive d'interrompre la formation, pourrait être d'autant plus conséquent que la charge de travail est élevée. Cette dernière hypothèse est d'autant plus plausible que nous avons constaté qu'il n'y avait pas de compensation de l'accroissement du nombre d'exercices par une diminution de la proportion des exercices exigeant a priori un investissement plus important, comme la réalisation de productions

évaluées par les pairs. Cette considération nous amène à revenir sur la question plus générale du lien entre caractéristiques des dispositifs et comportement des utilisateurs.

Les conjectures qui suivent, si elles dérivent en partie des analyses de structures présentées dans cet article, sont largement inspirées des travaux théoriques de Tinto (1975), mais aussi des réflexions sur l'*effet test* (Roediger et Karpicke, 2006). Nous suggérons que les exercices automatisés ont joué un rôle central dans la persistance des utilisateurs au sein des MOOC, en influant sur l'un des paramètres majeurs du modèle de Tinto (1975), la performance académique (Grade Performance). En d'autres termes, si l'on se rapporte au modèle de Tinto, les apprenants qui ne réalisent pas ces exercices, par choix ou parce que le dispositif n'en propose pas ou peu, voient l'un des mécanismes majeurs de leur persistance au sein du dispositif sensiblement amputé. En favorisant l'obtention de notes élevées par des exercices automatisés au niveau d'exigence relativement faible, les concepteurs ont pu jouer, sciemment ou non, sur le paramètre qu'est l'intégration académique, et partant de là, sur l'attachement au but (Goal Commitment), i.e., la volonté d'obtenir le certificat.

Ceci constitue une différence majeure par rapport aux Open Coursewares (OCW) (D'Oliveira *et al.*, 2010), qui ne présentent pas, pour la plupart, d'exercices interactifs associés au cours. En plus du rôle sur la persistance, ces exercices ont probablement favorisé la mémorisation du contenu (Agarwal *et al.*, 2008) par l'effet test qu'ils induisaient, y compris dans le cas, le plus fréquent, où le contenu du MOOC était accessible aux apprenants au moment de l'évaluation. Néanmoins, une étude comparative entre MOOC et OCW, qui sort du cadre de cet article, serait nécessaire pour donner à ces conjectures un ancrage empirique.

## **Conclusion**

L'étude présentée ici a considéré un nombre important de cours avec un niveau de détail qui n'aurait pas été permis par une analyse manuelle, mais elle présente un certain nombre de limites. En premier lieu, il reste à tester empiriquement le lien entre comportement des utilisateurs et structure du cours, notamment via des analyses de traces d'interaction. En second lieu, les typologies mobilisées ici pour classifier les items sont celles qu'imposent les plates-formes sur lesquelles sont hébergés les cours. Nous avons fait ici abstraction des modèles taxonomiques qui permettent de guider l'élaboration de ces items, et qui pourraient notamment nous permettre de mieux en appréhender la fonction. Ce sont des limites inhérentes à l'analyse par un programme *ad hoc*. Celui-ci nous permet d'avoir un aperçu immédiat sur plusieurs dizaines de milliers d'exercices, mais seule une analyse didactique plus fine, et qui *a minima* imposerait une analyse semi-automatisée, nous permettrait d'aborder la question de la fonction de ces exercices.

Se pose enfin la question de l'étayage de plusieurs des hypothèses que nous avons soulevées au cours de la discussion. Sur la base de nos résultats, nous émettons notamment l'hypothèse selon laquelle, dans la mesure où le concepteur est conscient du fait que le certificat n'est pas formellement reconnu, il attribue généralement d'autres fonctions à l'évaluation que celles qui dominent généralement dans les cours universitaires, comme ce peut être le cas de la validation d'acquis : s'assurer que les vidéos produites ont bien été visionnées, permettre aux apprenants de s'autoévaluer pour mieux retenir les principaux messages du cours. Il conviendrait pour corroborer une telle hypothèse de se pencher davantage sur les visées que les concepteurs de MOOC donnent à ces cours, les valeurs qu'ils colportent.

Pour conclure sur la question des perspectives de recherche, nous suggérons, pour de futurs travaux, de consacrer davantage d'attention à la question de l'influence des politiques des plates-formes sur les caractéristiques des dispositifs. Depuis 2016, sur des plates-formes comme Coursera (Koller, 2015), les utilisateurs ne payant pas n'ont plus accès qu'aux seules vidéos pédagogiques et aux rares exercices optionnels pour l'obtention du certificat.

Cette nouvelle donne est susceptible d'inciter les concepteurs à revoir à la hausse le niveau d'exigence des évaluations pour donner à ces certificats une valeur plus adaptée aux investissements financiers consentis par les apprenants. Les mutations actuelles soulèvent nombre de questions quant à l'influence réciproque des stratégies des institutions à l'origine des cours, et des stratégies des plates-formes hébergeant les MOOC. Il reste en particulier à déterminer l'impact de la disparition progressive des exercices gratuits sur la persistance au sein des MOOC pour les utilisateurs ne payant pas, mais aussi sur les modalités d'évaluation. De ce point de vue, l'utilisation de programmes analogues à celui que nous avons mobilisé ici pourrait s'avérer féconde.

La centration de la communauté scientifique sur les utilisateurs des cours, à travers moult analyses de traces d'interaction ou de questionnaires, conduit sans doute à négliger les recherches sur les caractéristiques des dispositifs et les tenants de leurs évolutions. Il y a fort à craindre, en conséquence, que l'on échoue à capturer la dynamique d'un écosystème de cours qui, s'il n'a pas encore révolutionné l'enseignement supérieur autant que certains le craignaient ou le prédisaient, a tout de même revivifié le débat sur le rôle que les institutions d'enseignement supérieur ont à jouer dans l'apprentissage en ligne.

## References

Agarwal, P. K., Karpicke, J. D., Kang, S. H. K., Roediger, H. L., & McDermott, K. B. (2008). Examining the testing effect with open- and closed-book tests. *Applied Cognitive Psychology, 22*(7), 861–876.

- Alraimi, K. M., Zo, H., & Ciganek, A. P. (2015). Understanding the MOOCs continuance: The role of openness and reputation. *Computers & Education, 80*, 28–38.
- Bean, J. P., & Metzner, B. S. (1985). A Conceptual Model of Nontraditional Undergraduate Student Attrition. *Review of Educational Research, 55*(4), 485–540.
- Bourdages, L., & Delmotte, C. (2007). La persistance aux études universitaires à distance. *International Journal of E-Learning & Distance Education, 16*(2), 23–36.
- Brown, K.M. (1996). The role of internal and external factors in the discontinuation of off-campus students. *Distance Education, 17*(1), 44-71.
- Cisel (2016). Utilisation des MOOC : éléments de typologie. Manuscrit de thèse non publié. Récupéré du site : <http://www.theses.fr/2016SACLN024>
- Daniel, J. (2012). Making Sense of MOOCs: Musings in a Maze of Myth, Paradox and Possibility. *Journal of Interactive Media in Education, 3*(0).
- D'Oliveira, C., Carson, S., James, K., & Lazarus, J. (2010). MIT OpenCourseWare: Unlocking Knowledge, Empowering Minds. *Science, 329*(5991), 525–526.
- Engle, D., Mankoff, C., & Carbrejy, J. (2015). Coursera's introductory human physiology course: Factors that characterize successful completion of a MOOC. *The International Review of Research in Open and Distributed Learning, 16*(2).
- Evans, S., & Myrick, J. G. (2015). How MOOC instructors view the pedagogy and purposes of massive open online courses. *Distance Education, 36*(3), 295–311.
- Greene, J. A., Oswald, C. A. et Pomerantz, J. (2015). Predictors of retention and achievement in a massive open online course. *American Educational Research Journal, 52*(5), 925-955.
- Henderikx, M. A., Kreijns, K., & Kalz, M. (2017). Refining success and dropout in massive open online courses based on the intention–behavior gap. *Distance Education, 38*(3), 353–368.
- Ho, A. D., Reich, J., Nesterko, S. O., Seaton, D. T., Mullaney, T., Waldo, J. et Chuang, I. (2014). *HarvardX and MITx: The first year of open online courses, Fall 2012-Summer 2013* (SSRN Scholarly Paper No. ID 2381263). Récupéré du site SSRN : <http://ssrn.com>
- Hollands, F. M., & Tirthali, D. (2014). Resource requirements and costs of developing and delivering MOOCs. *The International Review of Research in Open and Distance Learning, 15*(5).
- Hollands, F. M., & Tirthali, D. (2015). *MOOCs in Higher Education: Institutional Goals and Paths Forward*. Palgrave Macmillan.
- Joo, Y. J., Lim, K. Y., & Kim, J. (2013). Locus of control, self-efficacy, and task value as predictors of learning outcome in an online university context. *Computers & Education, 62*, 149–158.
- Jordan, K. (2014). Initial trends in enrollment and completion of massive open online courses. *The International Review of Research in Open and Distributed Learning, 15*(1).
- Jordan, K. (2015). Massive open online course completion rates revisited: Assessment, length and attrition. *The International Review of Research in Open and Distributed Learning, 16*(3).

- Kemp, W. C. (2002). Persistence of Adult Learners in Distance Education. *American Journal of Distance Education*, 16(2), 65–81.
- Khaneboubi, M., Baron, G. (2015). Analyse de la participation à un cours en ligne ouvert à tous : le cas d'« enseigner et former avec le numérique (eFAN) ». *Revue internationale des technologies en pédagogie universitaire*, 12(1-2)
- Kizilcec, R. F., Pérez-Sanagustín, M., & Maldonado, J. J. (2017). Self-regulated learning strategies predict learner behavior and goal attainment in Massive Open Online Courses. *Computers & Education*, 104(Supplement C), 18–33.
- Koller, D., Ng, A., Do, C., & Chen, Z. (2013). Retention and intention in massive open online courses. EDUCAUSE Review.
- Koller, D. (2015) An update on assessments, grades, and certification. Récupéré à <http://Coursera.tumblr.com/post/131520811622/an-update-on-assessments-grades-an>
- Kolowich, S. (2013). The Professors behind the MOOC Hype. *Chronicle of Higher Education*. Récupéré à l'adresse <http://chronicle.com/article/The-Professors-Behind-the-MOOC/137905/>
- Kulkarni, C., Wei, K. P., Le, H., Chia, D., Papadopoulos, K., Cheng, J., ... Klemmer, S. R. (2013). Peer and Self Assessment in Massive Online Classes. *ACM Trans. Comput.-Hum. Interact.*, 20(6), 33:1
- Lee, Y., & Choi, J. (2011). A Review of Online Course Dropout Research: Implications for Practice and Future Research. *Educational Technology Research and Development*, 59(5), 593–618.
- Ma, J., Han, X., Yang, J., & Cheng, J. (2015). Examining the necessary condition for engagement in an online learning environment based on learning analytics approach: The role of the instructor. *The Internet and Higher Education*, 24, 26–34.
- Margaryan, A., Bianco, M., & Littlejohn, A. (2015). Instructional quality of Massive Open Online Courses (MOOCs). *Computers & Education*, 80, 77–83.
- Radford, A. W., Robles, J., Cataylo, S., Horn, L., Thornton, J., & Whitfield, K. E. (2014). The employer potential of MOOCs: A mixed-methods study of human resource professionals' thinking on MOOCs. *The International Review of Research in Open and Distributed Learning*, 15(5).
- Roediger, I., & Karpicke, J. (2006). Test-Enhanced Learning: Taking Memory Tests Improves Long-Term Retention. *Psychological Science*, 17(3), 249–255.
- Rootman, I. (1972). Voluntary Withdrawal from a Total Adult Socializing Organization: A Model. *Sociology of Education*, 45(3), 258–270.
- Rovai, A. P. (2003). In Search of Higher Persistence Rates in Distance Education Online Programs. *Internet and Higher Education*, 6(1), 1–16.
- Spady, W. G. (1970). Dropouts from higher education: An interdisciplinary review and synthesis. *Interchange*, 1(1), 64–85.
- Thomas, S. L. (2000). Ties That Bind: A Social Network Approach to Understanding Student Integration and Persistence. *The Journal of Higher Education*, 71(5), 591–615.

- Tinto, V. (1975). Dropout from Higher Education: A Theoretical Synthesis of Recent Research. *Review of Educational Research*, 45(1), 89–125.
- Tinto, V. (2006). Research and Practice of Student Retention: What Next? *Journal of College Student Retention: Research, Theory et Practice*, 8(1), 1–19.
- Toven-Lindsey, B., Rhoads, R. A., & Lozano, J. B. (2015). Virtually unlimited classrooms: Pedagogical practices in massive open online courses. *The Internet and Higher Education*, 24, 1–12.
- Zaromb, F. M., & Roediger, H. L. (2010). The testing effect in free recall is associated with enhanced organizational processes. *Memory & Cognition*, 38(8), 995–1008.

## Auteur

**Matthieu Cisel.** Laboratoire STEF, Post-doctoral student, ENS Paris-Saclay. E-mail: matthieucisel@gmail.com

---

<sup>i</sup> Précisons dès à présent que l’attrition désigne, dans les paragraphes qui suivent, la part des personnes qui s’inscrivent à un diplôme sans l’obtenir ; la rétention et la persistance représentent l’inverse de l’attrition, à savoir la part de ceux qui obtiennent le diplôme. Dans le contexte d’un MOOC, ces termes ont une signification sensiblement différente, puisque l’on s’intéresse à l’obtention du certificat d’un MOOC, par ailleurs gratuit, et non d’un diplôme nécessitant plusieurs années d’investissement.

<sup>ii</sup> Une API est une interface de programmation applicative ; c’est une interface fournie par la plate-forme d’hébergement à partir de laquelle des informations peuvent être extraites, dans notre cas au format JSON. Le JSON est un format de données textuelles, dérivé de la notation des objets du langage *Javascript*.

<sup>iii</sup> Précisons que FUN est une plate-forme dont les inscriptions sont gratuites, et sur la période analysée ici, qui s’étend de janvier 2014 à juin 2015, les certificats étaient tous gratuits.

<sup>iv</sup> Nous n’avons néanmoins pas cherché à quantifier l’étendue d’outils autres que ceux de la plate-forme FUN.