HEC MONTRÉAL

L'influence du format vidéo sur l'engagement et l'apprentissage dans le contexte de cours en ligne : une perspective neuroscientifique

par

Sergej Lackmann

Sciences de la gestion

(Option Intelligence d'affaires)

Mémoire présenté en vue de l'obtention

du grade de maîtrise ès sciences

(M.Sc.)

Septembre 2019

Sommaire

De plus en plus d'institutions éducationnelles offrent des cours en ligne et des millions d'étudiants suivent ceux-ci. La grande majorité de ces cours utilise la vidéo comme outil principal d'enseignement. Ainsi, il devient très important de comprendre quel effet l'utilisation de la vidéo a sur l'étudiant. L'objectif de ce mémoire est de comparer deux types de formats de vidéos pédagogiques en fonction de l'engagement émotionnel et cognitif qu'ils créent chez les étudiants à travers le temps. En plus, nous voulons savoir quel format facilite le plus l'apprentissage des étudiants et mieux comprendre la nature de la relation qui existe entre l'engagement et la performance des étudiants dans le contexte des cours en ligne. L'engagement émotionnel et cognitif des étudiants est recueilli à la fois par les questionnaires (c.-à-d. explicite) et les mesures neurophysiologiques (c.-à-d. implicites). En particulier, nous utilisons les données implicites (l'électroencéphalographie, l'activité électrodermale, et les expressions faciales) pour mieux comprendre les états émotifs et cognitifs, souvent inconscients, des sujets.

Pour répondre à nos questions de recherche, nous effectuons une expérience laboratoire intersujet avec deux conditions comprenant un seul facteur : le type de vidéo. Dans la première condition, la vidéo est de type « enregistrement de cours magistral » et dans la deuxième condition elle est de type « enrichi ». Dans ce contexte, « enrichi » signifie que la vidéo a une meilleure présentation visuelle, ce qui permet de mieux transmettre le contenu aux spectateurs. Les deux vidéos ont le même contenu et la même longueur. Au total, 26 sujets ont été assignés d'une manière aléatoire à l'une des deux conditions.

Les résultats suggèrent que la vidéo enrichie maintient un engagement émotionnel (activation) et cognitif plus élevé à travers le temps que l'enregistrement de cours magistral. Cependant, la valence émotionnelle à travers le temps est plus positive parmi les étudiants qui regardent le cours magistral. En ce qui concerne l'apprentissage, la vidéo enrichie permet une meilleure maîtrise des questions difficiles. En fin de compte, il existe un lien significatif entre l'engagement cognitif et émotionnel des étudiants et la performance de ceux-ci.

Notre recherche démontre l'intérêt d'utiliser les mesures neurophysiologiques pour les recherches futures sur les formats de cours en ligne. Ces mesures permettent de recueillir des données implicites à travers le temps et de saisir la dynamique temporelle d'engagement. Dans le cas de notre étude, ces données permettent de mieux comprendre l'expérience d'apprentissage au-delà les réponses aux questionnaires, ainsi que de discerner les différences entre les conditions en matière d'engagement et les liens entre engagement et performance.

Mots-clés : engagement, MOOC, neuroscience, vidéo pédagogique, apprentissage

Table des matières

Liste des tableaux et des figures

Liste des figures

Liste des tableaux

Avant-propos

Le mémoire suivant est présenté sous la forme d’un article avec l’accord de la Direction du programme de la M.Sc., HEC Montréal. Le consentement a été obtenu de la part des auteurs pour présenter cet article dans le contexte de ce mémoire. De plus, le CER a approuvé le projet de recherche qui a servi à produire cet article.

Remerciements

Bien que le mémoire porte mon nom, plusieurs personnes ont contribué directement ou indirectement à sa réalisation.

J'aimerais donc remercier profondément mon directeur de mémoire, Pierre-Majorique Léger, qui m'a permis de vivre cette expérience de recherche. Je lui serais toujours reconnaissant pour son encouragement et son soutien continu. Je suis également très reconnaissant à mon codirecteur Patrick Charland qui m'a soutenu lorsque j'ai dû faire face à une courbe d'apprentissage élevée. C'est grâce à lui que j'ai pu approfondir mes connaissances en neuroscience et en éducation. Je tiens aussi à remercier l'équipe de recherche, Carolin Aube et Jean Talbot, pour leur temps et pour leur engagement envers ce projet et ma réussite.

Il aurait été difficile d'effectuer ce travail sans le soutien de l'équipe du Tech3Lab. J'aimerais remercier Élyse Labonté-Lemoyne en particulier pour son mentorat et soutien technique au niveau des outils et données neurophysiologiques. Je suis aussi redevable à Carl St-Pierre qui m'a permis d'amener ma maitrise de l'analyse de données à un tout autre niveau.

Enfin, j'aimerais souligner le support de ma famille et de mes amis qui ont été à mon côté tout au long de cette aventure.

Chapitre 1: Problématique et questions de recherche

Mise en contexte de l'étude

En 2016, environ 58 millions des personnes ont suivi des cours en ligne de type ouvert et massif, communément appelés « MOOC » (d'après l'acronyme anglophone Massive Open Online Course) et ce nombre est en forte augmentation (Allen, 2015; Shah, 2016). Ces cours sont souvent offerts d'une manière asynchrone et leur contenu est majoritairement composé de vidéos qui exposent le matériel pédagogique (Guo, 2014, Hansch, 2015). Ces vidéos varient beaucoup en format et en longueur, car les créateurs possèdent et utilisent différents types de matériel pour les produire. Les chercheurs de ce domaine (Chen, 2015; Da Silva, 2016; Guo, 2014; Hansch, 2015; Kizilcec, 2015) argumentent que le format choisi a un impact immédiat sur l'expérience des personnes qui suivent ces cours, ainsi que sur les coûts de développement d'un MOOC. Alors, l'un des volets de la recherche s'intéresse à l'engagement et à l'apprentissage en lien avec ces différents formats de vidéo et au choix du meilleur format.

En effet, des études ont suggéré que les divers formats ont des effets différents sur l'apprentissage et l'engagement des étudiants. Par exemple, Lee (2014) affirme qu'une vidéo qui contient seulement des diapositives engage plus les étudiants et permet un meilleur apprentissage par rapport à une vidéo qui inclut aussi l'avatar d'un enseignant. Selon le chercheur, si une vidéo inclut trop d'images et est trop complexe, l'apprentissage des étudiants diminue.

La littérature de ce domaine, notamment Chen (2011), appelle à poursuivre la recherche afin de comprendre l'impact des formats sur l'apprentissage et l'engagement dans différents contextes, par exemple en changeant des facteurs comme le sujet de la vidéo. Par ailleurs, plusieurs études (Chen, 2015; Wang, 2008, Wang, 2017) soutiennent que l'utilisation des outils neurophysiologiques et l'analyse des mesures implicites ouvrent sur de nouvelles possibilités de comparaison en matière d'engagement des vidéos. En particulier, il sera intéressant d'analyser l'évolution d'engagement à travers le temps (Dillon, 2016).

Questions de recherche

Notre étude vise globalement à comparer l'effet du format des vidéos sur l'engagement et l'apprentissage des apprenants. Spécifiquement, nous allons évaluer l'engagement généré par une vidéo de type « enregistrement de cours magistral » et une de type « enrichi ». La vidéo de type « enrichi » est caractérisée par une présentation audiovisuelle qui inclut des images, des graphiques et du texte en lien avec la narration qui explique le contenu d'apprentissage.

Nous voulons ainsi identifier lequel des deux formats mène à un meilleur engagement et s'il produit davantage d'apprentissages chez les apprenants. Nous tenons compte de l'évolution d'engagement à travers le temps et mesurons deux composants de l'engagement, soit l'engagement émotionnel et cognitif. Ainsi, notre première question de recherche (RQ1) est la suivante : Quel format évoque un meilleur engagement chez les étudiants à travers le temps? Et la deuxième question de recherche (RQ2) est celle-ci : Quel format est susceptible d'offrir un meilleur apprentissage chez les étudiants?

Enfin, nous cherchons à examiner davantage la relation entre l'engagement et la performance des étudiants. En particulier, nous analysons les liens entre l'engagement émotionnel et cognitif mesurés de manière implicite et explicite, et l'apprentissage des étudiants. La littérature existante propose qu'il y a une corrélation entre les deux (Chen, 2015). Alors, nous voudrions comprendre si ce lien peut être validé dans le contexte de notre étude et déterminer le comportement de cette relation. La troisième question de recherche (RQ3) est la suivante : Quel lien existe-t-il entre l'engagement des étudiants et leur performance?

Objectifs de l'étude et contributions potentielles

Au niveau théorique, notre étude vise à contribuer à la littérature existante. D'abord en comparant la vidéo de type enrichie qui inclut des graphiques, des animations et d'autres caractéristiques visuelles stimulantes. Ce format n'avait pas été analysé précédemment. Ensuite, notre étude vise à explorer davantage l'engagement émotionnel et cognitif des étudiants à travers les mesures implicites neurophysiologiques correspondantes. Cette recherche sera la première à comparer les mesures implicites neurophysiologiques et explicites (questionnaire) liées à l'engagement émotionnel et cognitif durant le visionnement de capsules vidéo de type magistral ou enrichi. Ainsi, nous pourrons mieux comprendre l'expérience vécue et son effet sur l'apprentissage des participants.

Au niveau pratique, notre objectif est de mieux comprendre la dynamique de l'engagement cognitif et émotionnel à travers le temps grâce aux mesures implicites saisies tout de long du visionnement des vidéos. Cela va permettre d'éclairer la dimension temporelle, car elle n'est pas encore bien comprise. Il sera alors possible de valider l'effet des deux formats sur les étudiants à travers le temps et de potentiellement émettre des balises de conception en matière de longueur des capsules.

De plus, l'étude permet de valider et de donner de meilleurs indices en ce qui concerne la relation qui existe entre l'engagement cognitif et émotionnel (implicite, explicite) et la performance des étudiants. En fin de compte, nous pourrons alors faire une recommandation quant au choix parmi ces deux types de formats et de futures considérations pour la conception des capsules.

Informations sur l'article

Les premières discussions portant sur le projet ont débuté à l'été 2016. Après ces discussions et la révision de la littérature actuelle, l'envergure et le protocole pour l'étude ont été finalisés à l'automne 2016. Afin d'obtenir le matériel pour les deux conditions, nous avons filmé une vidéo de type « enregistrement de cours magistral » à la fin de l'année 2016. La capsule de type « enrichi » existait déjà. Ensuite, au début de 2017, nous avons effectué un prétest pour valider notre approche et la faisabilité du projet.

L'expérience a été réalisée au printemps 2017 au Tech3Lab de HEC Montréal. Les résultats préliminaires de cette collecte ont été présentés lors la conférence Open edX 2018. Aussi, un résumé a été publié dans les actes de la conférence EduLearn en juillet 2018. L'article de ce mémoire sera soumis au journal « Computers in Human Behavior ».

Résumé de l'article

Les cours en ligne attirent des millions des personnes. La majorité de ces cours utilisent la vidéo comme outil pédagogique principal. Plusieurs formats sont utilisés et la littérature actuelle soutient que chaque format a des effets différents en matière d'engagement et d'apprentissage. Nous comparons deux formats : une vidéo de type « enrichi » avec une de type « enregistrement de cours magistral ». En particulier, nous voulons savoir quelle vidéo engage plus les étudiants et est propice à un meilleur apprentissage. Aussi, nous analysons s'il existe un lien entre l'engagement des étudiants et la performance de ceux-ci.

Pour répondre à ces questions de recherche, nous avons mené une expérience auprès de 26 personnes. Chaque personne a regardé une des deux vidéos et nous avons recueilli des mesures neurophysiologiques (l'électroencéphalographie (EEG), l'activité électrodermale (EDA), et les expressions faciales) à travers l'expérience. Au début et à la fin de l'expérience, les sujets ont répondu à un questionnaire qui portait sur le contenu de la capsule pour mesurer l'apprentissage. L'analyse des données a été réalisée avec SAS (SAS Institute Inc., Cary, NC, USA) et NeuroRT Suite (Mensia Technologies, Paris, France). La théorie de la richesse des médias et la théorie cognitive sur l'apprentissage multimédia ont été mobilisées pour expliquer les résultats obtenus.

Nos données suggèrent qu'au début, les étudiants sont engagés émotionnellement plus avec la capsule non enrichie. Cependant, la capsule enrichie maintient mieux cet engagement à travers le temps. En ce qui concerne l'engagement cognitif, la vidéo enrichie soutient encore une fois mieux cet engagement à travers le temps et elle permet aussi aux étudiants de comprendre davantage les questions difficiles. En fin de compte, il existe une relation significative entre l'engagement des étudiants et la performance de ceux-ci.

Pour les futures études sur l'apprentissage en ligne, il sera donc intéressant d'utiliser aussi des mesures implicites neurophysiologiques. Nos résultats indiquent qu'ils permettent non seulement de mieux comprendre l'état de l'étudiant, mais ils corrèlent aussi avec l'apprentissage des apprenants. Étant donné que la capsule non enrichie évoque plus d'engagement émotionnel en début de visionnement, il sera intéressant de tester un mode qui pourrait combiner les avantages de ces deux formats.

Contributions et responsabilités personnelles

Activité	Contribution
Revue de littérature	Revue de la littérature sur l'apprentissage (en ligne), l'engagement et les outils neurophysiologiques – 100 % Repère et mobilisation des concepts/théories du domaine, ainsi que justification de la pertinence des questions de recherche.
Création du design expérimental	Création du protocole de l'expérience –90 % Le design expérimental a été valide par des prétests. Par la suite, la séquence des tâches, le choix des outils neurophysiologiques et des stimulus ont été concrétisés et reflétés dans le protocole. Développement du stimulus (vidéos)– 25 % Les requis pour la production ont été développés par l'équipe de recherche et l'équipe opérationnelle a élaboré les stimulus. Création de la demande au CER et des demandes de changement – 90 % À partir des modèles, nous devions créer ces demandes et développer des formulaires de consentement et de compensation. David Brieugne a grandement soutenu mes efforts.

	Mise en place des places de collecte (équipement) – 10 % La majorité du travail a été effectuée par l'équipe opérationnelle.
Recrutement des sujets	Élaboration du questionnaire de recrutement – 100 % Recrutement des sujets – 60 % Gestion de compensation et suivi des sujets – 90 %
Prétests et collecte	Chargé des opérations lors des collectes – 90 % Responsable de la résolution de problèmes techniques et fonctionnels avec l'équipe opérationnelle
Extraction et transformation des données	Extraction et transformation des données – 100 % Avant de pouvoir faire les analyses, il fallait extraire les données des différents outils neurophysiologiques et les transformer d'une manière à ce que celles-ci puissent être analysées à l'aide des autres programmes.
Analyse de données	Nettoyage et préparation des données – 100 % Analyse statistique – 90 % Le nettoyage et la préparation des données ont pris considérablement d'effort et de temps. Le mentorat et le coaching de Carl St-Pierre étaient essentiels pour accomplir ces tâches.

Rédaction	Rédaction de l’article et du mémoire – 100 % À travers de nombreux commentaires et conseils, mes directeurs Pierre-Majorique Léger et Patrick Charland m’ont aidé à réaliser ce travail.

Chapitre 2: Article

Using neuroscience to evaluate the influence of video production styles on engagement and learning in MOOC context

Sergej Lackmann, HEC Montréal, Montréal, Canada

sergej.lackmann@hec.ca

Pierre-Majorique Léger Ph.D., HEC Montréal, Montréal, Canada

pierre-majorique.leger@hec.ca

Patrick Charland, Ph.D., Université du Québec à Montréal, Montréal, Canada

charland.patrick@uqam.ca

Abstract

Millions of students follow massive open online courses (MOOC). Most of these MOOCs are delivered in video format and therefore the video production style is crucial for cost, student engagement, and learning results. There are a growing number of studies that examine video production styles and their impact on engagement and learning. However, most of these studies focus on self-perceived and behavioural measures. In the research field, there is still a need to investigate the cognitive and emotional state of learners using neurophysiological measures (which provide the evolution over time), link those measures to the theoretical foundations of engagement and compare all prevalent production styles in the MOOC field.

Our study aims to compare the lecture capture and the animated video production style, in terms of engagement over time and learning, in a between-subject lab experiment. We use Fredrick's multidimensional definition of engagement as the theoretical link between video production style and learning. Engagement is operationalized using implicit neurophysiological data (electroencephalography, electrodermal activity, facial expressions) and perceived (i.e., self-reported) emotional valence and activation. In total, 26 subjects participated in the study and were randomly assigned to one of two conditions. In both conditions, the participants watched a 15-minute lecture video on an introduction to psychology. In the first condition, the video format was lecture capture (Lecture condition), and in the second video, we used the animated production style (Animated condition).

The results show that lecture capture engages emotionally subjects more over a shorter period, while the animated production style maintains higher emotional and cognitive engagement over longer periods of time. As for student learning, the animated production style allows for significantly improved performance pertaining to difficult questions. Additionally, our results suggest that there is a significant relationship between engagement and student performance. In general, the higher the engagement, the better the student performance, though, in the case of cognitive engagement the link is quadratic (inverted U shaped).

Building on our results, it seems that the animated production style is better suited for longer online lectures, and lecture capture for shorter ones. A combination of the two could be the best option regardless of the length and it is justifiable to invest in high-quality videos to increase student engagement, while incorporating social cues. Furthermore, it is relevant to analyze production styles and examine student engagement (and its evolution over time) using implicit neurophysiological data as it gives more insight than do explicit measures.

Keywords: MOOC, engagement, video format, learning, neuroscience

1. Introduction

Massive open online courses (MOOCs) are classes that are characterized by a large number of students who can simultaneously follow those classes online without real-time interaction with an instructor (Hansch, 2015). The classes are usually provided by online learning platforms such as Coursera (https://www.coursera.org/), Edx (https://www.edx.org/), or Udacity (https://www.udacity.com/) which allow users to access the content through a website.

According to Shah (2016), in 2016 the total number of students in MOOCs is estimated at 58 million. Since 2013, 74% of all new higher-education students opt for online learning (Allen, 2015). However, engagement of students over time is an issue and retention rates are often below 10% (Veletsianos, 2016; Xiong, 2015). Given its growing importance as a new phenomenon, MOOCs are a salient research topic in education (Ilioudi, 2013; Kizilcec, 2015; Wang, 2017).

Most MOOCs are primarily a compilation of videos that expose the learning content (Guo, 2014; Hansch, 2015). Content is not necessarily created by the online learning platform, but by content creators such as universities. Due to differences in practices and resources available to those content creators, videos are produced in a variety of ways and the formats vary substantially within a platform (Hansch, 2015).

At the same time, researchers and practitioners in the field argue that choosing the right video format for a MOOC is crucial for student engagement, cost, and learning results (Chen, 2015; Da Silva, 2016; Guo, 2014; Hansch, 2015; Kizilcec, 2015). Several studies suggest that some video production styles have a larger impact on learning than others. On one hand, videos that include an instructor's image result in higher student performance (Kizilcec, 2015; Ilioudi, 2013; Wang, 2017). On the other hand, video designs that are too complex can overload student's cognitive capacity and decrease learning (Lee, 2014; Wang, 2017). It is therefore necessary to analyze the impact of video lecture elements on learning performance and use those findings to improve the format (Chen, 2015).

As for cost, Hansch (2015) suggests that creating the video is the most expensive item in producing a MOOC. In particular, to create a lecture video you need content, filming equipment, a production site, and often post-editing. The process can become complex and time consuming (Chen, 2015; Guo, 2014). Guedes (2016) suggests that beyond the learning efficiency, the choice of the appropriate video design should be cost-effective.

As for engagement and learning, in traditional school settings, instructional material influences learner engagement (Ogbu in Fredricks, 2004). Engagement, it is argued, lowers drop-out probability (Ekstrom in Fredricks, 2004), and increases learning (Hew, 2016). Recent studies indicate that this relationship between instructional material (video production style in particular), engagement, student retention and learning holds true in the online learning environment. Therefore, video format becomes an important element of analysis. First, based on previous research, Chen (2011) argues that design elements of multimedia learning materials have an impact on emotions and those emotions correlate with subject performance. This view is also supported by Hansch (2015) in his research paper who argues that video production quality is important to keep a student's attention (cognitive engagement). Other studies suggest that some video formats are better than others in soliciting a student's engagement and satisfaction (Lee, 2014).

As engagement is argued to be a strong predictor of retention in MOOCs (Xiong, 2015), and retention rates are often below 10% (Veletsianos, 2016; Xiong, 2015), the choice of the best video production styles could achieve higher engagement, and consequently lower dropout rates in MOOCs. Da Silva (2016) outlines that the "negative effect of possible discrepancy between enrolment and completion noticed … can be overcome with good practices and with carefully designed videos".

Given the dominant role of video teaching in MOOCs, and the impact of video elements on learning, as well as engagement, it is relevant to analyze different video production styles in order to achieve better pedagogical objectives and higher levels of retention in the online environment (Chen, 2015; Ilioudi, 2013). At the same time, several authors suggest that more research is necessary in this field (Chen, 2011; Hansch, 2015; Ilioudi, 2013; Sun, 2007).

Even though there are a number of studies that examine video production styles and their impact on engagement and learning (Guo, 2014), "survey data and secondary data collected via automated methods dominated the analyses" in the MOOC field (Veletsianos, 2016, p. 17). That data includes user-generated activity such as video viewing times automatically captured on MOOC servers. Yet, it was demonstrated that neurophysiological measures allow us to better understand students' experience in e-learning (Chen, 2011; Harley, 2015). Engagement evolves throughout time and is subject to important retrospective bias. Therefore, there is a benefit to using these implicit measures to capture automatic and unconscious reactions of subjects (de Guinea, 2014; Wang, 2008). No study until now has compared the lecture capture and the animation video production styles by analyzing the evolution of engagement over time (the duration of the video) (Dillon, 2016). Furthermore, engagement has been defined differently by researchers in the MOOC field which warrants clarification of the construct and its operationalization. Current research also seeks to further examine the relationship between student engagement and performance in different contexts (Chen, 2015).

We aim to advance current findings by comparing the lecture capture and animated production styles in terms of which one engages the students more emotionally and cognitively (Research Question 1) over time (the duration of the video), and which one provides better learning outcomes (Research Question 2). Also, we further examine the relationship between student engagement and learning outcomes (Research Question 3). Using implicit and explicit measures to infer a subject's engagement, a between-subject experiment with 2 conditions is used to answer the research questions. In the Animated condition, the students watch the animated production-style video showing animated graphics, images and text. In the Lecture condition a lecture capture is shown, which was filmed in a class setting where the professor and students were visible. Both videos are approximately 15 minutes long.

2. Related work

Current literature discusses the video production styles using the media richness theory and the cognitive theory of multimedia learning (CTML) (Chen, 2015; Homer, 2008; Kizilcec, 2015). These theoretical frameworks help to outline the conceptual differences between the two video production styles. The differences have specific consequences in terms of learning outcomes and engagement. We use the Hansch (2015) definition of video production style which outlines the concept as "the main method of visual organization that is employed to realize a video's goals" (Hansch, 2015, p. 20).

2.1 Cognitive Theory of Multimedia Learning

Mayer (2005) developed the CTML which aims to explain the relationship between multimedia learning and the cognitive processes of students regarding the assimilation of knowledge. This model outlines three predictions. First, a person needs to be actively processing incoming information for learning to happen. This means that a person needs first to actively use his senses (hearing, seeing) to perceive the information so that it can then be sent to and processed in the working memory (Clark & Mayer in Kizilcec, 2015).

Second, during “learning”, information is gathered and processed in the student’s working memory. Visual and auditory information is treated separately, in two channels (Baddeley, 2003; Homer, 2008; Kizilcec, 2015). The visuospatial sketchpad processes the visual information, such as an image on the screen, and the phonological loop handles auditory stimuli, such as the voice of the professor. Third, at any given time, the visuospatial sketchpad and phonological loop can handle a constrained amount of information in relation to the limited capacity of the working memory. After the information has been processed in the working memory, it is integrated with previous knowledge in the long-term memory.

As for the first prediction, several authors in the multimedia learning field argue that actively attending to the visual and verbal stimuli of the multimedia lecture is required for effective learning (Kizilcec, 2015; Korving, 2016). In videos, they suggest that the social cues (voice, face) of the lecturer would increase the student’s engagement and therefore would lead to better performance (Guo, 2014; Homer, 2008; Korving, 2016). Korving (2016) outlines that a person’s face naturally draws attention to it, and that due to social customs people react in a more engaged manner. Furthermore, in accordance with the social response theory, if students see a professor’s image on the screen, they consider the computer to be a social actor and believe they are in a human-to-human interaction (Nass in Lee, 2014). Video lectures that feature the instructor’s image supposedly make it easier for students to pay attention, improve student engagement, satisfaction and perceived learning more than those that do not (Kizilcec, 2015; Wang, 2017).

Consistent with the dual channel principle, video and audio input delivered at the same time is suggested to provide better results in terms of engagement and learning performance in empirical studies (Chen, 2011). In particular, seeing the image of a professor and hearing the narration is suggested to provide for better learning since the image of the instructor offers additional nonverbal cues compared to using PowerPoint slides and narration (Wang, 2017).

The instructor's lips, facial expressions and gestures could provide additional information in the form of nonverbal cues which activate social interaction schemas and help students better grasp the material being taught. While visuals of the professor are processed by the visuospatial sketchpad, the narration is handled by the phonological loop. Both information sources do not interfere with each other and are even complementary (Korving, 2016; Kizilcec, 2015; Wang, 2017).

However, Lee (2014) points out that only PowerPoint slides compared to slides with a human-like animated character elicit the most arousal and learning outcomes. Also, he found that arousal is correlated with performance. He argues that students are used to that type of teaching and therefore their body (arousal) reacts more to that type of stimulus as students immerse themselves in their learning routine. Also, several authors suggest that the professor's image in an instructional video can create additional cognitive load (compared to narration only) and burden the students which would offset the benefits of the social cues and nonverbal communication (Kizilcec, 2015, Wang, 2017). Too much load on the visuospatial sketchpad was found to limit the ability of students to attend to all the content presented visually in multimedia learning, thus inhibiting the student's acquisition of knowledge (Chen, 2015; Homer, 2008; Wang, 2017), and also negatively impact their engagement (attention, emotions) (Chen, 2011, 2015; Wang, 2017). This could be especially the case if the instructor's image does not provide additional information which is already presented on the slides or through narration, or even takes attention away from additional information on the slides (Homer, 2008). Homer's (2008) results show no difference in terms of learning and social perception between an instructional video with a professor and PowerPoint slides compared to only PowerPoint slides and narration.

Other results suggest that there exists a temporal dynamic of attention as in Korving (2016). The visibility of the lecturer with PowerPoint slides allows students to pay easier attention and benefit from nonverbal communication only after 15 minutes of continuous watching. It's only after that time that students seem to report significantly higher attention rates while seeing a large image of the professor and PowerPoint slides compared to only PowerPoint slides. After 15 minutes, the students' attentional resources are depleted and they start to pay more attention to the professor's image as it conveys nonverbal cues and helps them understand the material compared to only PowerPoint slides. Korving (2016) argues that in the first 15 minutes, the students' attentional resources allow them to maximize their learning objectives and to read the slides, as well as listen to the narration. This does elicit more sustained attention in the first 15 minutes compared to the professor's image with PowerPoint slides.

In summary, current literature suggests that visual and auditory stimuli in instructional videos need to be pertinent in terms of improving engagement and learning, as well as providing a positive trade-off for the cognitive load they can cause. Instructor presence does not necessarily elicit better engagement, satisfaction and perceived learning. The benefits of social presence and nonverbal communication can be counterbalanced by more demand being placed on the visual channel due to the processing of social cues. The trade-off seems to be less prominent when students are lively and worsens the more students are exposed to the stimuli. Further research can clarify these benefits and trade-offs of different visual and audio elements.

2.2 Media richness theory

According to Trevino, Lengel, & Daft (1987) communication channels have different levels of media richness. Level of richness in this context is referred to as how much a medium can efficiently transmit the information that allows people to have a prompt and common understanding. Thus, some communication means are better in transmitting information than others and allow for a rapid reduction in uncertainty and equivocality to produce the same interpretation.

Four criteria allow us to determine the level of richness of a medium. First, the faster the medium allows for feedback, the better it is at reducing misunderstanding. Second, the more cues the medium provides (images, graphics, numbers, voice, body language), the better it is at facilitating the right interpretation. Third, a diverse vocabulary which includes numbers and natural language provides rigour (numbers) as well as context (natural language) at the same time and improves communication. Fourth, a medium which includes personal focus, as well as emotional responses will provide additional personal meaning and increase the efficiency of the communication (Trevino, 1987; Sun, 2007). A medium which excels better at those four criteria would then be richer than a medium which does not meet all the criteria. The theoretical propositions for multimedia learning would be that instructional material with higher levels of richness according to the four criteria will be better at transmitting information and conveying meaning. Consequently, the students would be able to learn faster and better with rich media compared to less rich media.

Empirical research suggests that rich media reduces the effort needed by parties to understand each other (Chen, 2015). In multimedia learning, Chang & Chang (2004) demonstrated that increasing the richness of the learning material with animation makes communication more efficient and leads to positive responses (better engagement) from students. In addition, Chen's (2011) study suggests that "video-based multimedia material can be a powerful learning tool that provides learners with a rich and rewarding experience" (Chen, 2011, p. 254). Compared to less rich non-video formats in the form of PowerPoint presentations and animations of the content, lecture capture provided the best emotional response and learning. Similar results were also reported by Ilioudi (2013) who found that lecture video with an instructor image was specifically better for learning complex topics compared to studying with a book. Also, the authors found that the talking head video was significantly better for knowledge acquisition than the khan-style video. In line with the fourth criteria of the media richness theory, the insertion of instructor image in video lectures is suggested to increase student engagement and learning due to arguably increased personal focus of that media and additional cues in the form of body language (Guo, 2014; Kizilcec, 2015; Wang, 2017).

Moreover, the findings of Moreno (2001) suggest that adding personal pronouns such as ‘you’ to the video lecture increases the cognitive engagement and learning of students. However, when Chen (2015) compared three different video lecture types which have seemingly different levels of richness, he found that students’ positive and negative emotional responses were similar. The research in this field can therefore benefit from further empirical studies into the effect of media richness on engagement and learning.

2.3 Learners’ Engagement

Many studies outline the importance of engagement in education and assert its impact on a student’s performance, satisfaction and retention. In a traditional school setting, engagement has long been acknowledged to have a positive impact on student performance (Aber in Fredricks, 2004; Hew, 2016), satisfaction (Fredricks, 2004), and dropout rates (Ekstrom in Fredricks, 2004). Recent studies outline the similar impact of engagement in online learning environments. It is argued that it improves student learning (Guo, 2004) and retention (Li, 2012). In particular, students’ emotions (emotional engagement) are linked to retention [in MOOCs] (Dillon, 2016) and learning (Harley, 2015). In the context of this research, it was shown that emotions (emotional engagement) can be influenced by the visual characteristics of multimedia learning material and in turn impact learning outcomes (Chen, 2011).

It needs to be stated that engagement has been defined and operationalized differently by researchers in the MOOC field which warrants clarification of the construct and its operationalization to avoid engagement being “everything to everybody” (Fredricks, 2004). Fredricks (2004) proposes a multidimensional definition of engagement which is composed of behavioural engagement, emotional engagement and cognitive engagement.

He supports the hypothesis that it is necessary to study those together, as they are interlinked and provide together a richer characterization of a student's state. Also, the level of engagement can vary, and higher levels lead to better results in terms of satisfaction and learning. Therefore, there is a need to measure differences in engagement intensity. Finally, "a multifaceted approach to engagement argues for exploring how attempts to alter context influence all three types and determining whether outcomes are mediated by changes in one or more components" (Fredricks, 2004, p. 4).

Behavioural engagement is divided in three dimensions: conduct, work involvement, and participation (Fredricks, 2004). Work involvement is defined as the effort of simply doing the work. Participation means to be involved in extracurricular activities, to contribute to class discussions, to ask questions (Birch in Fredricks, 2004). In the case of a MOOC, it would be to attempt quizzes after a video or to participate in forum discussions. Conduct is positive behaviour, such as showing up for class, or not spamming in forums. However, even if people are at the task, it does not mean they are actively thinking about it.

Emotional engagement is broadly defined as positive or negative reactions (Fredricks, 2004) that do not require cognitive effort and that are spontaneous (Wang, 2008). Those reactions can be discrete emotions such as interest, boredom, happiness, sadness and anxiety (Connell in Fredricks, 2004). In turn, those discrete emotions can be grouped under arousal (i.e., activation) and valence (Russell in Harley, 2015). Charland (2015) defines valence as a pleasant or unpleasant emotional state (e.g., happiness, sadness), while arousal describes physiologically aroused or calm state (e.g., anxiety, boredom). Different scales and models can be then used to represent qualitative differences in valence and pleasure (Wang, 2008), and deduct a level of emotional engagement.

A person is cognitively engaged when actively trying to understand new information (Fredricks, 2004). In particular, cognitive engagement is a "psychological process involving attention and investment" (Marks in Smiley, 2011, p. 18). For Fredricks (2004) sustained cognitive engagement requires cognitive effort (processing of information) and focus during learning over time.

To learn, we first need to focus, maintain our attention on the stimulus and block out other irrelevant information so as not to overload our cognitive system (Driver in Chen, 2015). Memorizing is an example of that process (Fredericks, 2004). In the MOOC context, it means not just watching a video and daydreaming, but paying attention to the stimuli and actually thinking about its content. According to Chen (2015), we need to examine how distinct video production styles affect sustained attention in e-learning, as it is a significant factor for learning.

3. Hypothesis development

Our general research objective is to analyze two different video formats (lecture capture and animated production style) in terms of engagement and learning. Images that move quickly in and out of the screen represent, in our case, animation (Rieber in Chen, 2011). Hansch (2015) outlined a typology of video production styles which we use to categorize our videos.

On the one hand, it is relevant to compare the lecture capture production style, because simple recordings of class lectures are prevalent in online teaching (Chen, 2015; Homer, 2008) and also in blended learning (Lagerstrom, 2015). On the other hand, it is relevant to analyze the animated production style because it has not been studied in previous research. Also, there are a high number of different production styles being used and it is relevant to understand their strengths and weaknesses relating to online learning. There is not necessarily one best format (Ilioudi, 2013) and not much research has been done to analyze all those different formats (Chen, 2015). It still needs to be understood how production value affects learning (Hansch, 2015) and comparing lecture capture to animated production style helps us to answer that question.

Our production styles are characterized by social cues, cognitive demand and media richness. Those characteristics have an impact on the emotional and cognitive engagement of students, as supported by our literature review. Emotional engagement can be subdivided into arousal and valence which represents different aspects of emotion and together provide a better understanding of the emotional state of subjects (Harley, 2015). We do not investigate behavioural engagement as students are monitored and know that they will be evaluated. As outlined in the previous chapter, engagement has an influence on learning performance. We measure learning by calculating how much better students answered post-test multiple-choice questions on the content of the videos compared to pre-test.

We use the previously explained media richness and the CTML to explain why emotional engagement (arousal, valence) is likely to be higher for the Animated condition. Current literature indicates that richer media according to the four criteria leads to better learning performance and higher engagement than less rich multimedia (Chen, 2011). We can argue that our videos have a different media richness, and that this difference in media richness between our two videos will produce a significant difference in emotional engagement. If we look at our videos, the animated production style shows animated context-rich images, has text and also audio. In this regard, it provides more visual cues (second criteria) and richer vocabulary (third criteria) than the lecture capture, which has only simple video and audio. Ultimately, the feedback (first criteria) given to students is also higher. This will produce more excitement (arousal) and satisfaction (valence) for the Animated condition.

Based on the previous discussion of the CTML, literature indicates that students are more emotionally engaged with familiar PowerPoint presentation-style formats (Lee, 2014), like the video in our Animated condition. Also, several studies indicate that the professor's image does not provide significantly more emotional engagement than other production styles which do not include it (Homer, 2008; Chen, 2015).

Human presence can activate substantial cognitive resources, increase the effort required to process the information and offset the benefits provided by social cueing (Chen 2015; Wang, 2017). In conclusion, in light of the CTML and media richness theory, we propose that emotional engagement (arousal, valence) will be higher for the Animated condition.

H1A: Emotional engagement will be higher for animated production style

As we discussed for emotional engagement, we argue that the rich content of the animated production-style video allows students to better sustain their attentional engagement compared to only focusing on the teacher, similar to the results of Korving's (2016) research. Maintaining focus on a single static point can be challenging. Thus, dynamic content with short shot sequences can be better at maintaining a student's attention (Da Silva, 2016). Chen's (2011) results also suggest that richer media provides a more immersive experience to students. The objects which appear on the screen such as tables and images are well-organized in the Animated condition. There is no second window which shows the professor or agenda. Therefore, the learner can focus all his attention on the information displayed without having to divide it between for example the professor's image and the text. In conclusion, we argue that cognitive engagement will be higher for the subjects in the Animated condition.

H1B: Cognitive engagement will be higher for animated production style

First, most research supports the idea that higher media richness has a positive impact on learning performance. As discussed previously, the animated production style has a higher media richness than lecture capture. Interesting visual effects and graphics which convey further meaning and provide more context are likely to support more learning, especially in terms of difficult questions.

Second, according to the CTML, multimedia material which provides audio and video stimuli will allow for better learning. On the one hand, an animated production-style video presents visual information in the form of graphics which is processed by the visual loop. At the same time, the narration provides additional explanations which is processed by the audio loop of the learner. On the other hand, the lecture capture provides narration and an image of a teacher in the class. However, a simple image of the teacher does not necessarily convey or support additional information processing in the video loop (Homer, 2008). It also seems that students could learn more with formats similar to animated production style (Lee, 2014). Therefore, we propose that animated production style will have a better impact on learning.

H2: Learning performance will be higher for animated production style

As discussed, many studies argue that there is a significant relationship between the emotional engagement of students and their learning (Chen, 2011; Harley, 2015; Homer, 2008; Lee, 2014). In particular, Lee's (2014) research shows that the socialness perception, arousal, and pleasure have an effect on a student's performance. When students see a familiar type of video lecture, their body reacts subconsciously to it and the students become alert (aroused) in order to prepare for instruction. This allows learners to be further engaged with the video lecture and consequently process more information.

While testing different video course designs, Chen (2011) found that there is a link between pre-test results and negative emotions with students' learning. The negative emotions had a negative impact on performance. When a student is upset or anxious, he does not necessarily focus his attention on studying and his mind is preoccupied. Inversely, social cues in the form of a human voice or image can have a positive impact on learning, as they can provide a student with a familiar feeling and decrease stress or anxiety. In summary, we expect a positive correlation between positive emotional states and performance. The higher the emotional engagement, the better the student performance.

H3A: The higher the emotional engagement, the better the student performance

Cognitive engagement and, more specifically, attention, is considered key for effective learning in multimedia learning (Korving, 2016; Serrhini, 2017; Steinmayr in Chen, 2011). Chen (2015) further proposes that video production style is an important focus of analysis as it could affect sustained attention. He argues that too high levels of sustained attention and its deviation during specific video types causes stress and as a result, a student's performance decreases compared to other video production styles. Using EEG data, he discusses in another study the negative correlation between students learning (post-test and progressive score) and low-attention intervals during video lectures (Chen, 2011), meaning that when students were not paying attention, they were performing worst. Inversely, we expect a positive correlation between higher sustained attention level and performance. However, this relationship is likely to be quadratic (inverted U shaped). At the beginning, more attention increases performance. At some point, however, too much attention leads to an overload of the visuospatial sketchpad. This negatively affects learning since the student cannot process any more information and begins to stress, as discussed in Chen's (2015) study. Finally, we propose the following hypothesis:

H3B: Quadratic relationship between cognitive engagement and performance

4. Material and Methods

To test our hypothesis, we conducted a between subject lab experiment with two conditions. In both conditions the subjects watched an approximately 15 minutes long pedagogical video on *Introduction to Organisational Behavior*. In the Animated condition the video production style was animated, in the Lecture condition it was a lecture capture. During viewing, emotional and cognitive measures were recorded with neurophysiological instruments. Furthermore, a pre- and post-test multiple-choice questionnaires were administered to measure student learning. Upon completion of the experiment, subjects were given a 30$ gift card for the university bookstore. A research protocol was created and approved by Research Ethics Board of our institution before the experiment has taken place.

4.1 Research Participants

16 male and 10 female subjects have participated in the study. They were recruited using panel of research participants from our institution. Subject were randomly assigned to one of two conditions and were pre-screened before participation. First, only subjects who had previously no psychology classes at the university level were admitted. This was done to assure that no one had prior knowledge of the content which was shown in the videos. Based on previous research, Homer (2008) outlined in his article that due to the expertise reversal effect, elements of instructional design do not have the same effect on learners who have previous knowledge compared to novices. Also, subjects could not have some sort of head covering and needed to tie back long hair to avoid shadows on or/obstruction of the face for automatic facial expression recognition software to function properly. In addition, subjects who had a mental condition (epilepsy, neurological surgeries, etc.) were excluded as this can cause deviations in the EEG signal (Charland, 2018).

4.2 Experimental Stimulus

Two lecture videos with the assistance of the same professor have been created specifically for the purpose of the study. Both videos explain exactly the same content; a chapter of the introductory psychology course at our institution and have approximately the same length (15 minutes). This is within the length range suggested for MOOC videos by previous research (Chen, 2015; Korving, 2016; Lagerstrom, 2015). In the Animated condition, learners see an animated production style video (according to Hansch's (2015) typology of video formats). Continuous flow of images, graphics and text is synchronized with the content being explained via an audio track. In the Lecture condition, subjects see a video recording of a class lecture (lecture capture). As suggested by recent research (Guo, 2014), filming of the lecture capture was planned, so it can be used in a MOOC. You can see only the professor and some students who attend the class. No other content such as PowerPoint slides is visible. The voice of the professor is the same in both conditions.

Figure 1: Animated condition screenshot

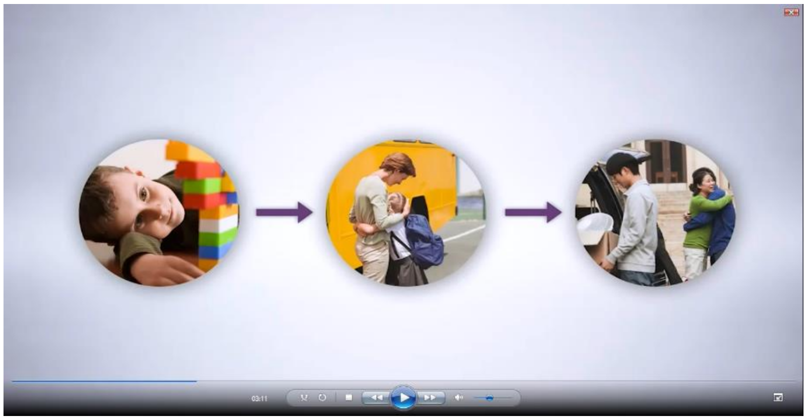


Figure 2: Lecture condition screenshot

Like in Chen (2015), the next table summarizes the characteristics of our video production styles in terms of the theory discussed: media richness and CTML.

Table 1: Research stimuli in comparison

Comparison item	Lecture capture production style	Animated video production style
Cost	Low	High
Conveyed learning context	Professor presenting the subject in a traditional class setting	A PowerPoint like presentation which shows dynamic content with a background voice of a professor
Multimedia elements	Camera focused on the professor, audio	Graphics, images, text, audio, flash animations
Media richness	Medium	High
Social cues	Many	Some

4.3 Measures

We infer students' emotional and cognitive engagement over time based on the related neurophysiological states (valence, arousal, attention) which we measure using implicit and explicit data. The data is collected by electroencephalography (EEG), electrodermal activity (EDA), facial expression classification software, and questionnaires. In particular, valence is operationalized through the analysis of facial expressions. Arousal is operationalized through EDA and facial expressions analysis. Self-reporting data is collected using a Self-Assessment Manikin (SAM) (Bradley, 1994) questionnaire which provides insights on explicit valence (pleasure) and arousal (anxiety). Cognitive engagement is understood under the concept of vigilance or attention (Mikulka, 2002). Attention is operationalized using EEG. Results can be cross validated when we aggregate those different measures to understand response to stimuli (Stewart in Wang, 2008). Building on the psychophysiology framework (Bagozzi in Wang, 2008), the below table illustrates the relation between our measures, the related neurophysiological states, and how we asses the physiological consequences (bodily responses) of those states.

Table 2: Operationalization of the measures

Measure	**Neurophysiological state (response to stimuli)**	**Operationalization**
Cognitive engagement	Cognitive response: attention	EEG, Brainvision (Morrisville, USA)
Emotional engagement	Affective response: valence	Facial expressions (FaceReader (Wageningen, Netherlands)), SAM pleasure
Emotional engagement	Affective response: arousal	Facial expressions (FaceReader (Wageningen, Netherlands)), EDA (Biopac (Goleta, USA)), SAM arousal
Learning performance		Difference between pre- and post-test multiple-choice questionnaire results

4.3.1 Emotional engagement

First, it needs to be noted that emotions are considered multi-componential and they can cause different bodily responses e.g. open mouth, increased heart rate etc. (Harley, 2015). These emotional components are divided into behavioral (facial), experimental (how emotions make one feel), and physiological (EDA) responses (Harley, 2015). Restated, "emotional response can be measured in at least three different systems - affective reports, physiological reactivity, and overt behavioral acts" (Lang in Bradley, 1994, p.49)

Affective reports allow to understand how participants perceive and feel about the stimuli (Harley, 2015). As for overt behavioral acts, Ekman (2000) outlined the link between facial expressions and six basic emotions (happiness, sadness, surprise, fear, disgust, and anger). Based on that work, several studies were able to successfully use automatic facial expressions analysis software to recognize emotions of participants in e-learning environment (Al-Awni, 2016; Bahreini, 2016; Lewinski, 2014).

Physiological reactivity in form of EDA can be measured by looking at the skin electrical conductance levels (SCL) and changes in those levels due to sympathetic activity and alter sweating. For that purpose, two electrodes pass little electricity through the skin and SCL is measured in micro Siemens or uS. EDA (arousal) is high when people are curious or anxious, and low then they are bored/relaxed (Harley, 2015). Electrodermal activity is widely and reliably used to measure arousal due to stimuli (Wang, 2008). In particular, Charland (2015) and Harley (2015) used EDA to assess students' arousal states during learning exercises.

Due to multi-componential nature of emotions, Harley (2015) argues that a multimodal approach (facial recognition, EDA, self-reporting) provides better effectiveness in measuring emotional engagement in an online learning environment. His study shows that there is no clear correlation between three methods to measure emotions: facial expressions analysis, self-reporting, and EDA. Also, Thayer (1978) suggested that arousal can be measured by looking at brain activity, heart rate, as well as pupil dilation and all those provide a different facet of arousal. Therefore, we use several physiological measures to capture the different dimensions of emotion.

4.3.2 Cognitive engagement

Based on frequency and amplitude of the signal, as well as the spatial location of its origin, EEG is used to infer physiological states of the subjects (Charland, 2018). For instance, if a person is relaxed, alpha waves (8–12 Hz) amplitude increases in parietal and frontal cortex. If a person is attentive and is actively thinking, beta waves amplitude increases (Serrhini, 2017). Theta waves indicate a state of sleep and are often recorded in parietal and temporal regions (Chen, 2011). Furthermore, EEG allows to assess attention in subjects (Charland, 2018; Serrhini, 2017). Chen (2015) used an EEG headset to measure sustained attention. He was able to detect low-attention spans in students who were watching video lectures.

Pope, Bogart and Bartolome (1995) developed a cognitive engagement index (vigilance index) based on the EEG signal's spectral decomposition. This engagement index has been widely used in different contexts (Charland, 2015). In particular, they argue that higher beta activity is related to increased vigilance, while increased alpha and theta to lower. Subjects in a state of high vigilance can better respond to stimuli (Charland, 2018). Theoretically, Freeman (2004) proposes that cognitive engagement and vigilance are equivalent. And vigilance can also be described as cognitive sustained attention (Berka, 2007). We compute our cognitive engagement ratio (attention) according to a modified version of the vigilance index as proposed by Mikulka (2002).

4.3.3 Learning performance

On the Qualtrics website, subjects completed a multiple-choice questionnaire (25 questions) to asses their knowledge of the content of the videos, before (pre-test) and after viewing it (post-test). Similar to previous research, there was no time limit to complete the questionnaire, and no feedback was given to avoid effects on the post-test (Chen, 2015; Homer, 2008; Wang, 2008). Pre-test and post-test questions are the same, as in Chen (2011). Recall and transfer questions were included, as used in Wang (2017). Each question was answered either correctly and counted as 1, or 0 when answered incorrectly. For each student, a % of correctly answered questions was calculated for each difficulty rating and overall. Learning was calculated by subtracting pre- and post-test results.

To create the difficulty classification, the complexity of questions was rated by 6 experts using 4-point Likert scale (1=easy, 4=most difficult), as suggested by Cronan (2012) and similar to previous research (Wang, 2017). Easy questions are measuring low complexity knowledge, while medium and difficult questions are measuring high complexity knowledge. According to Bloom's revised taxonomy, students who can answer effectively difficult questions have acquired a deeper understanding of the material. They can analyse a new problem using acquired knowledge and apply that knowledge to solve it.

Questions which were answered wrong by =>50% of the experts were removed as ambiguous. Median difficulty of the questions based on expert rating is 2. Since our scale was from 1-4, the median difficulty is therefore representative of a medium difficulty. Questions were classified based on the median perceived complexity rating by the experts, see Cronan, (2012). Results are presented in the Table 3.

Table 3: Classification of questions

Median perceived complexity rating	**Difficulty classification**
6 questions with median = 3	high difficulty (24% of all questions)
5 questions with median = 2.5	medium difficulty (20% of all questions)
9 questions with median = 2	medium difficulty (36% of all questions)
5 questions with median = 1.5	easy difficulty (20% of all questions)

4.4 Apparatus

Noldus FaceReader (Wageningen, Netherlands) is used as automatic facial emotion recognition software, BrainVision ActiChamp 32 (Morrisville, USA) to capture EEG, and Biopac (Goleta, USA) to record EDA data. The method used for data acquisition is based on Charland's (2015, 2018) framework. Self-Assessment Manikin (SAM) questionnaire is employed to capture explicit perception of students of valence and arousal. Learning performance is measured by pre- and post-test multiple-choice questionnaires.

The SAM allows to measure pleasure, arousal, and dominance (control) using a non-verbal illustrated approach which can be used to assess a subject's affective state in different experimental conditions (Bradley, 1994). There is empirical evidence that heart rate, EDA, as well as facial emotions are linked to arousal and pleasure as reported with the SAM.

Whereas dominance is not as effective to determine an affective state as the other two (Bradley, 2014). In previous research, Lee (2014) used SAM to compare different video production styles. Consequently, we use SAM to capture self-reported arousal and valence of participants. After viewing the video, a web page opens which displays the SAM to the learners.

Noldus FaceReader software records subject's facial expressions and measures the intensity of Ekman's six basic emotions + neutral of the participants. In detail, the Active Appearance Model captures the facial expressions, while an artificial neural network classifies those, and computes a standardized valence and arousal value. The value reported ranges from -1 to 1 with 30 inferences per second. The higher the value the more the person is pleased or aroused. FaceReader has been validated and used in similar studies to ours (Charland, 2015; Harley, 2015; Hetland, 2016).

To capture electrodermal activity, a wireless amplifier (Biopac MP) and two electrodes placed on the palm of the non-dominant hand are used as apparatus, see Courtemanche (2017). Biopac has been successfully employed in a variety of studies to measure arousal (Courtemanche, 2017; Pauna, 2018). EEG data collection is performed with BrainVision. It's a non-invasive technique which uses electrodes placed on scalp surface to record EEG signal (Chen, 2011). We use the international system 10/20 for electrode deposition as proposed by American Encephalographic Society (1994).

4.5 Experimental protocol

During the experiment there was no break and a session lasted on average 1 hour and 30 minutes, including the setup and calibration of the equipment. Upon arrival, a subject needed first to read and sign the *Consent Agreement* and it was verified that the person corresponds to our selection criteria. Then, participants were told that they take part in a study which evaluates different lecture video designs and they were given a short summary of the experiment.

After the setup of equipment and with the start of the EEG baseline, all apparatus was recording continuously throughout the experiment until all tasks were completed. The data was synchronized using markers which delineate each part of the experiment in all the equipment (e.g. start / end of the video). All tasks were performed on a 22-inch flat screen in front of the participants. In detail, subjects were asked first to relax and close their eyes for 1m 30 sec in order to have a baseline for the EEG signal. This reduces possible anxiety and other subject specific neuropsychological deviations before the experiment starts and allows to have a more valid EEG sample. (Harmon-Jones in Charland, 2018).

Then, during one minute participants saw and counted randomly colored squares which appeared on the screen for 6 sec each. This serves as a “vanilla” baseline, same as in Courtemanche (2018). During the next step, subjects completed a multiple-choice questionnaire (25 questions) to asses their pre-test knowledge of the content explained in the video they were about to see. Similar to previous research, there was no time limit to complete the questionnaire, and no feedback was given to avoid effects on the post-test (Chen, 2015; Wang, 2017).

After completing the assessment, subject started to watch one of the two videos. No note taking nor pausing were allowed during viewing to assure the comparability of the neurophysiological data over time between the two conditions and subjects. Right after the video, a SAM questionnaire was administered. As a post-test, the same questions as in the pre-test were used to assess the learner's performance after viewing the video.

4.6 Data processing

After acquisition data is prepared, synchronized, and analyzed in SAS (SAS Institute Inc., Cary, NC, USA) using statistical models appropriate for each research question/data type. EDA data from Biopac, FaceReader’s valence and arousal, cognitive engagement ratio from Mensia, SAM results, and pre- and post-test answers are exported into .csv files and then imported to SAS.

We analyse neurophysiological data recorded during the time the subjects watched the lecture. Subjects who have too much missing or invalid data are removed. These issues can be due to, for example, electrode contact loss (EEG, EDA) or facial recognition failure (FaceReader). The statistical significance level of our hypothesises is p=0.05.

For neurophysiological data, we use a repeated-measures multiple linear regression with random effects (subject) to test if there is a significant difference (over time/on average) in arousal, valence, or attention between conditions, similar to Sanders (2016) and Charland (2018). We argue that neurophysiological data can be modeled using time series since fluctuations depend on past values. Thus, we use proc mixed with autoregressive covariance structure (see SAS doc). We include random-effects for the origin / intercept in order to account for variability between individuals. Method used is maximum likelihood.

Furthermore, we compute the performance for the pre- and post-test questionnaires and calculate how much students have learned. To verify if there are significant differences in learning between conditions, we perform a Signed rank test. Additionally, we use the Mann-Whitney U Test because of the small sample (Ilioudi, 2013) to perform our control analysis. The relationship between SAM's valence/arousal and performance is analysed using Spearman correlation coefficients given the small sample size.

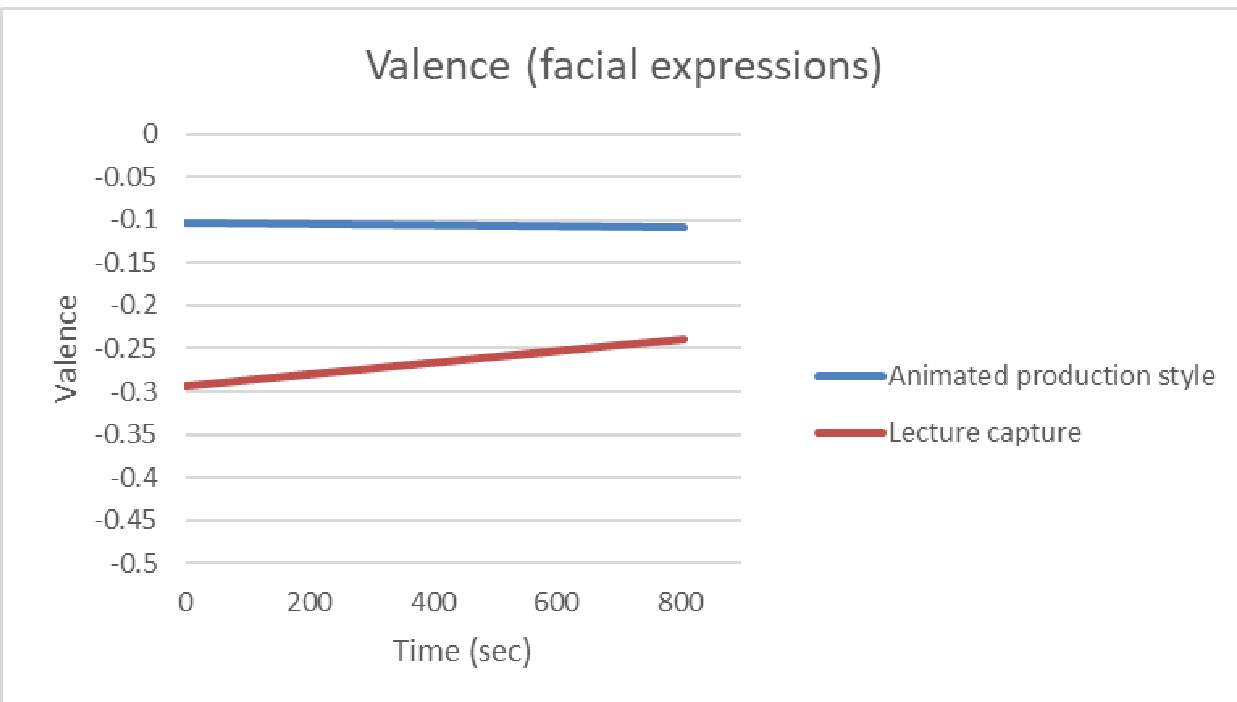
Each student has only one performance value (% of correctly answered questions) pre-test and post-test. Therefore, we first compute a mean value per subject for FaceReader's valence and arousal, arousal based on EDA, as well as attention (EEG). Then, we use multiple linear regression (proc reg) to model the relationship between the performance as dependent variable and a neurophysiological measure as independent variable. We also include quadratic relationships and condition as a control variable. This analysis is of exploratory nature given the number of subjects in each condition and the number of variables to include in the model. The annex provides further details as to how the data is analyzed and the results computed.

5. Results

5.1 Emotional engagement

Our results summarized in Table 4 suggest that there is no significant difference on average in valance between the conditions (β=−-0.19, Sig.=0.15>0.05) according to facial expression analysis. However, there is a significant difference in the evolution of valence over time (β= 7.30 x 10 ˆ -6, Sig.=<.0001<0.05), as can be seen in Figure 3. While valence for animated production style decreases over time, the valence of the students who watch the lecture capture video increases over time.

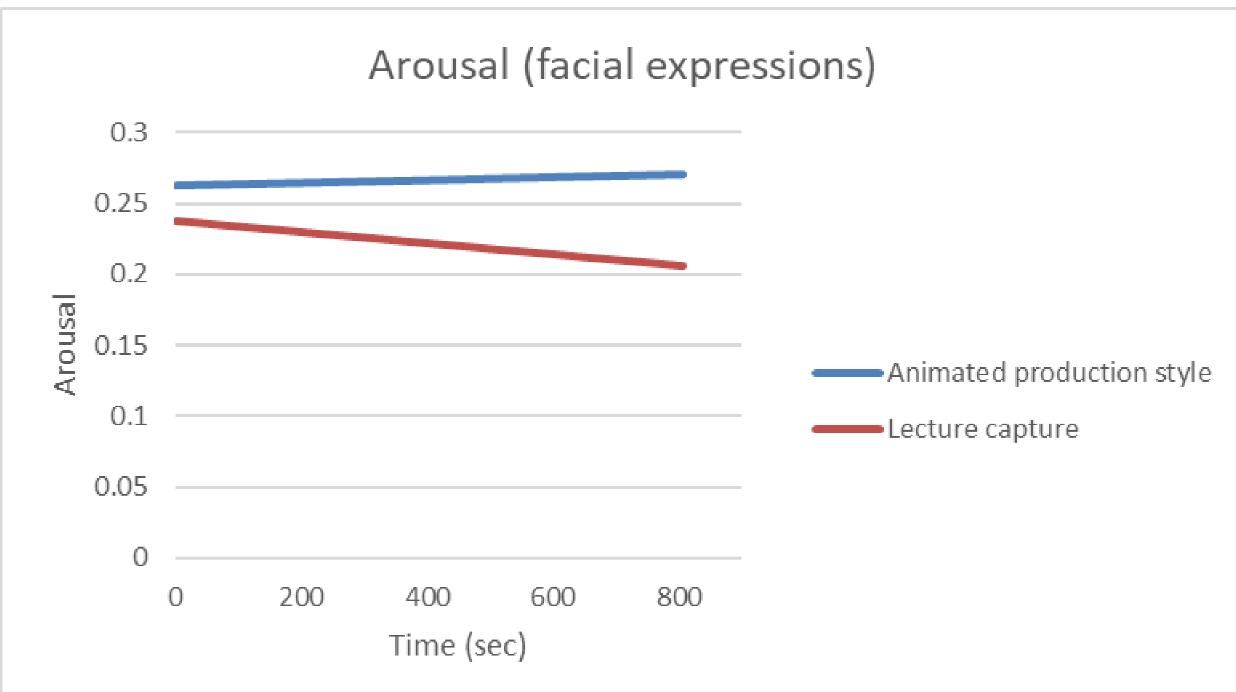
Figure 3: Evolution of valence over time based on facial expression analysis



Explicit valence as reported with SAM seems to confirm the above finding, as there is no significant difference in self-reported valence between the conditions (Sig.=0.24>0.05). In general, subjects in both conditions are however pleased with the videos: animation production style (Mean 6.17, STD 1.47) and lecture capture (Mean 6.86, STD 1.66).

There is no significant difference in arousal between the conditions (β=-0.02, Sig.=0.26>0.05) according to facial expressions analysis. The subjects in Animated condition have on average a similar arousal to the subjects in Lecture condition. However, there is a significant difference in the evolution of arousal over time (β=-5 x 10 ˆ -6, Sig.=<.0001<0.05). This is visualized in the Figure 4. While arousal for the lecture capture decreases over time, the arousal of the students who watch the animated production style increases over time.

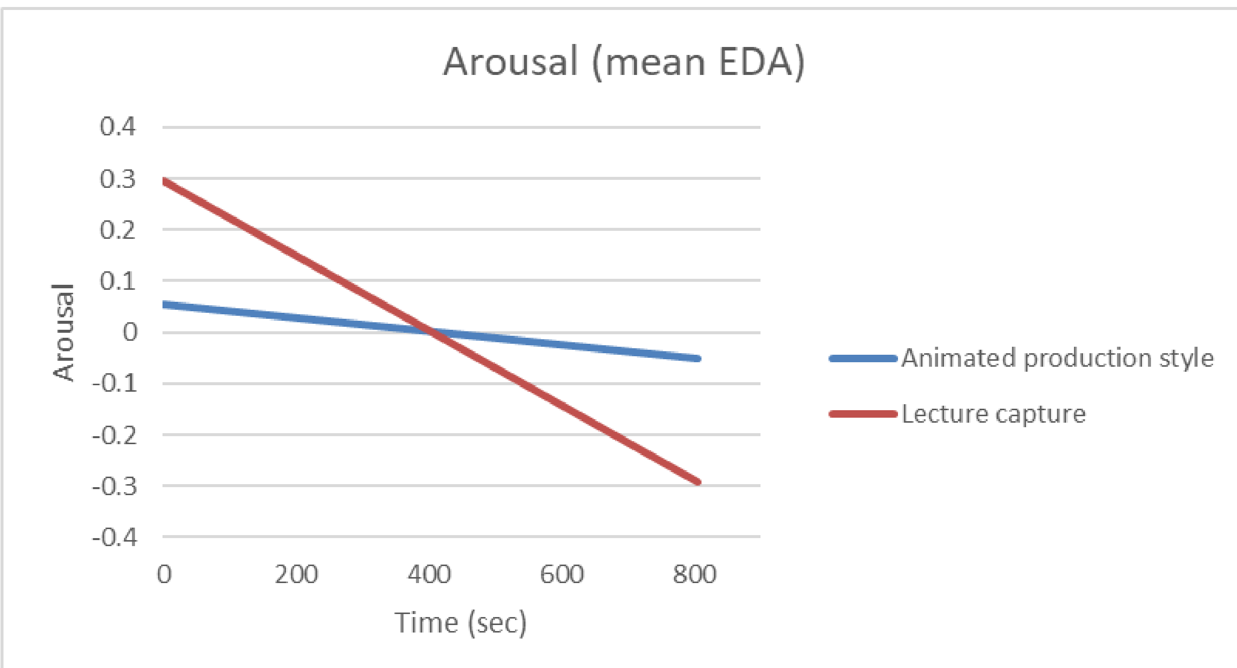
Figure 4: Evolution of arousal over time based on facial expression analysis



Self-reported arousal seems to confirm the above finding, as there is no significant difference between the conditions (Sig.=0.53>0.05). In general, subjects in both conditions were somewhat excited while watching both videos: animated production style (Mean 5.42, STD 1.62), lecture capture (Mean 4.79, STD 2.39).

As for arousal operationalized through mean EDA, there is a significant difference between the two conditions (β=0.24, Sig.=<.0001<0.05), as seen in Table 6. The subjects in the Animated condition have on average a lower mean arousal compared to the subjects in the Lecture condition. However, over time, the animated production style is able to incite significantly more arousal than the lecture capture (β=-60 x 10 ^ -6, Sig.=<.0001<0.05). At some point the former manages to invoke more arousal than the latter. It's clearly visible in the Figure 5 below, after both curves intersect.

Figure 5: Evolution of arousal over time based on mean EDA analysis

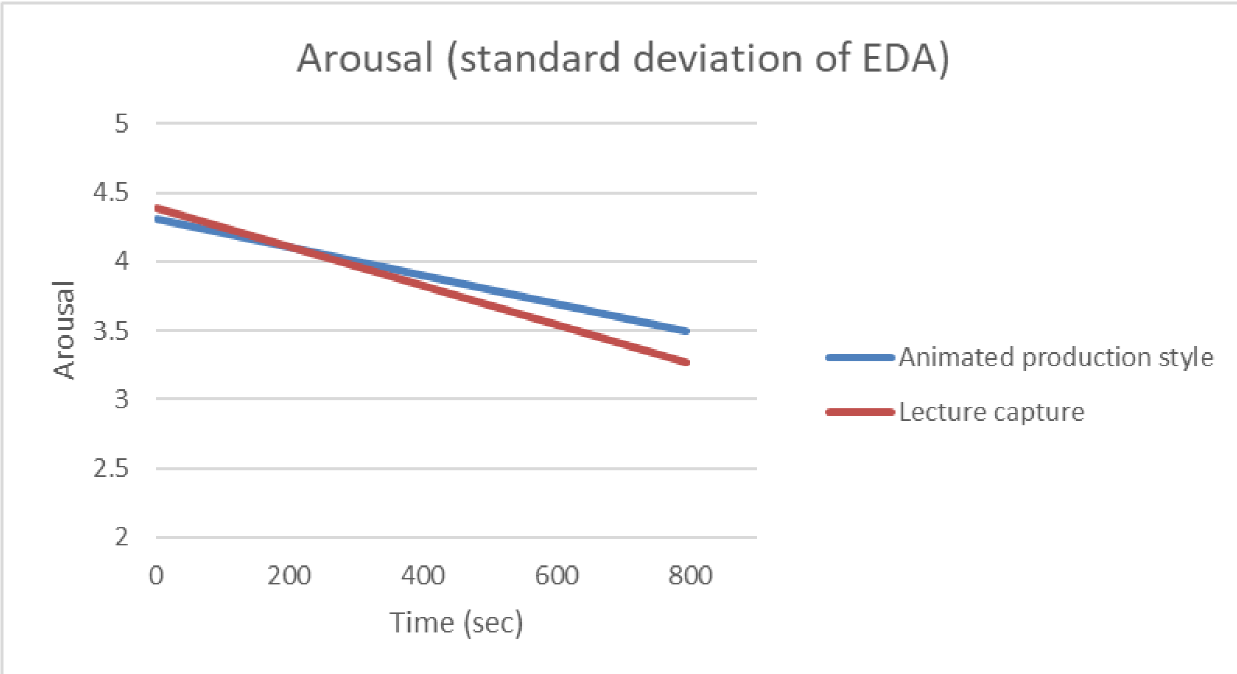


It is interesting to note that arousal in both conditions decreases over time, as one might think the students get tired. However, this has been not observed with the implicit arousal deducted from facial expressions. Also, the latter and explicit arousal did not exhibit a significant difference between conditions. Nevertheless, both implicit arousal measures (facial emotions and mean EDA) show that lecture capture induces less arousal over time compared to animated production style. Finally, we see an interesting spike at the beginning of Lecture condition in mean EDA.

As for deviation in implicit arousal according to EDA recordings, our results summarized in Table 7 suggest that there is no significant difference between the conditions (β=-0.004, Sig.=0.58>0.05). The subjects who watched the animated production style video have on average a similar arousal variability to the subjects who watched the lecture capture.

However, over time, the animated production style is able to incite significantly more arousal deviation than the lecture capture (β= 2.4 x 10 ^ -6, Sig.=0.02<0.05), as seen in Figure 6. And, at some point the former manages to produce more arousal variability than the latter. It's clearly visible in the graph below, as both curves intersect. This result is supported by implicit arousal operationalized using facial expressions and mean EDA. It is interesting to note that variability of EDA in both conditions decreases over time. We can assume the students get tired over time and their EDA variability stabilizes as they respond less and less aroused to external stimuli.

Figure 6: Evolution of arousal over time based on deviation of EDA analysis



As for emotional engagement, mean arousal according to EDA shows that the lecture capture engages significantly more the students on average. However, valence/arousal deducted from facial expressions, explicit arousal/valance, and deviation in arousal according to EDA do not show any significant differences between the conditions on average.

If we take time into account, facial valance increases significantly more over time for lecture capture, but facial arousal and arousal measured by EDA (mean and standard deviation) increase significantly more over time in the Animated condition (time coefficients are significantly different between conditions). In conclusion, the results do not support H1A: Emotional engagement will be higher for the animated production style.

5.2 Cognitive engagement

According to EEG recordings, our results suggest that there is no significant difference on average in attention between the conditions (β=-0.06, Sig.=0.81>0.05). The subjects in Animated condition have on average a similar attention level compared to the subjects in Lecture condition. However, there is a significant difference in the evolution of attention over time (β= -44 x 10 ˆ -6, Sig.=0.03<0.05), as seen in Figure 7. The attention of subjects in Animated condition is increasing, while in Lecture condition it's decreasing. Restated, even though attention in both conditions starts off similarly (significant intercept), over time the animated production style engages significantly more the students. Therefore, the results do support H1B: Cognitive engagement (attention) will be higher for the animated production style.

Figure 7: Evolution of attention over time based on EEG analysis

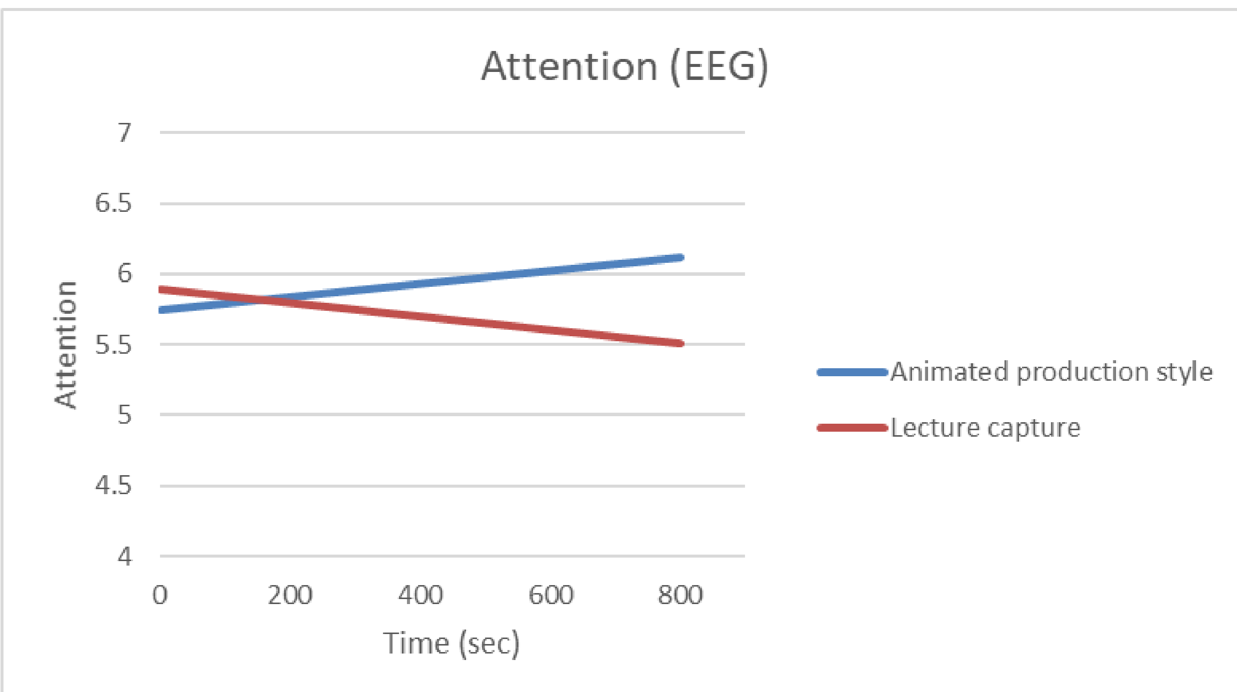


Table 4: Emotional engagement based on facial expression analysis

Variable	*Valence (mean)*	*Valence (over time)*	*Arousal (mean)*	*Arousal (over time)*
β – Animated condition (reference)	0	0	0	0
β – Lecture condition	-0.19	7.30 x 10 ˆ -6	-0.02	-5 x 10 ˆ -6
t Value	-1.45	4.97	-1.12	-10.73
Sig.	0.15	<.0001	0.26	<.0001

Table 5: Emotional engagement based on SAM analysis

Variable	*Valance*	*Arousal*
β – Animated condition (reference)	-	-
β – Lecture condition	-	-
t Value	-	-
Sig.	0.24	0.53

Table 6: Emotional engagement based on EDA analysis

Variable	*Arousal (mean)*	*Arousal (over time)*	*Arousal (STD)*	*Arousal (STD over time)*
β – Animated condition (reference)	0	0	0	0
β – Lecture condition	0.24	-60 x 10 ˆ -6	-0.004	2.4 x 10 ˆ -6
t Value	8.22	-9.48	-0.55	2.43
Sig.	<.0001	<.0001	0.58	0.02

Table 7: Cognitive engagement based on EEG analysis

Variable	*Attention (mean)*	*Attention (over time)*
β – Animated condition (reference)	0	0
β – Lecture condition	-0.06	-44 x 10 ˆ -6
t Value	-0.23	-2.11
Sig.	0.81	0.03

5.3 Learning outcomes

To control for bias, we do a control analysis. First, students should not be able to achieve pre-test more than 70% of correct answers for any questions type (easy, medium, difficult). Otherwise, our subjects have systematically previous knowledge of the content (experts). Overall, students performed pre-test similarly for all question types, except easy questions. These were answered correctly less than 50% of the time. It’s plausible that if someone does not know the topic, it’s difficult to answer easy questions as those are usually simply recall questions which are easy once you see the video. No question type is answered above 70% correctly. Therefore, we assume that the subjects don’t have systematically previous knowledge of the topic in the videos (experts). Also, participants in both conditions perform pre-test about the same (Sig.=0.11>0.05). We can therefore deduct that there is no bias in our conditions pre-test.

Post-test, students should perform significantly better after watching the video. Learning should have occurred in both conditions. This hypothesis is confirmed by our results. Students perform significantly better after watching the videos, regardless of the condition and question type (Sig.=0.04<0.05). In particular, easy and difficult questions saw the highest performance increase. This result supports the idea that once the subjects see the video it is easier to answer easy and difficult questions.

Comparing the two conditions, we see that participants in Animated condition are performing overall 3% better and their performance has increased 6% more. This overall difference between conditions, however, is not significant (Sig.=0.20>0.05). If we examine the different questions types separately, the results show that there is a significant difference in terms of mastering difficult questions (Sig.=0.04<0.05). The subjects in Animated condition have a performance increase twice of the other condition. The number of correctly answered difficult questions has increased by 38% in Animated condition and only by 18% in the Lecture condition. 4 out of 6 difficult questions are answered better by subjects of the Animated condition. The results support H2: Learning performance will be higher for animated production style.

5.4 Relationship between engagement and learning

Our hypothesis is that the higher the emotional and cognitive engagement, the better the student performance. As we have one directional hypothesis (positive effect), we divide p-values by 2. Furthermore, we need to take into the account that we have a small sample size and that certain variables included in the models are non significant, therefore the probability of F (goodness of the model fit) can be non significant. However, p-values retain their validity as we are interested in identifying the characteristic which impact performance (inference of variables).

5.4.1 Emotional engagement

Many psychologists and neurologists have outlined that emotions impact cognitive learning (Chen, 2011). Furthermore, in previous research about multimedia-based learning, it was shown that video production styles have an influence on emotions and that these emotions have consequences on performance (Chen, 2011). Students who were depressed, angry, or anxious had trouble learning (Goleman in Chen, 2011). So, valence can interfere with learning.

In our study, there is a significant positive relationship between overall performance and valance of the subject based on facial expression analysis. The higher the valance, the higher the performance (adjusted R=0.13, β=20.27, Sig.=0.03<0.05). That effect is true regardless the condition, as can be seen in Table 8. Similar results have been obtained by Chen's (2011) study who was looking into video-based multimedia learning. Students performance was significantly correlated with negative emotions when taking pretest scores into account (Chen, 2011). However, our self-reported valance seems to contradict the above finding, as there is no significant correlation between explicit valance and performance (Sig.=0.53>0.05).

Table 8: Multiple linear regression results for valance (facial expressions)

Variable	*β Estimate*	*t Value*	*Sig.*
Intercept	77.17	17.97	<.0001
Valence	20.27	2.25	0.03
Condition	2.49	0.43	0.67

Model fit: Adjusted R2=0.13, F-value=2.56, Sig.=0.10>0.05

As for arousal based on facial expressions analysis, there is a significant positive relationship between overall performance and mean arousal of the subject, if we include deviation of arousal into the model. The higher the mean and variability of arousal, the higher the performance (adjusted R=0.11, β=123.62, Sig.=0.03<0.05 (one-tailed) and β=110.22, Sig.=0.04<0.05 (one-tailed)). That effect is true regardless the condition, as can be seen in Table 9.

Table 9: Multiple linear regression results for arousal (facial expressions)

Variable	*β Estimate*	*t Value*	*Sig*
Intercept	13.92	0.52	0.61
Mean Arousal	123.62	2.01	0.06
STD Arousal	110.22	1.83	0.08
Condition	5.72	0.88	0.39

Model fit: Adjusted R2=0.11, F value=1.78, Sig.=0.19>0.05

According to arousal measured by EDA, our results suggest a significant positive relationship between mean arousal and performance, when including deviation of arousal into the model (same as for facial arousal). The higher the mean arousal, the higher the performance (β=110.22, Sig.=0.04<0.05). However, this effect is mediated by variability of arousal. As deviation of arousal increases, performance decreases (β=-16.19, Sig.=0.02<0.05). That effect is true regardless the condition. Table 10 provides the details of the results. Also, Lee's (2014) study shows that socialness perceptions, arousal, pleasure, flow experience, and learning motivation can all affect students' learning outcomes.

Table 10: Multiple linear regression results for arousal (EDA)

Variable	*β Estimate*	*t Value*	*Sig.*
Intercept	69.37	10.01	<.0001
Mean EDA	9.30	2.14	0.04
STD EDA	-16.19	-2.38	0.02
Condition	-1.70	-0.35	0.73

Model fit: Adjusted R2=0.15, F value=2.21, Sig.=0.12>0.05

It is important to note that this significant relationship is only observable if we calculate mean EDA for the first 5 minutes of the viewing time. There is no significant relationship between mean EDA and performance if we take the average of the whole video session (15 min). This result can be explained by the evolution of arousal over time (RQ1). Over the duration of the video, mean EDA decreases significantly for both conditions (β= -13 x 10 ^ -6, Sig.=0.005<0.05). Subjects get tired, mean EDA is lower at the end of the video viewing for both conditions, regardless the performance of the students. This masks the relationship between high mean EDA and high performance overall.

Explicit arousal reported through SAM partially confirms the above results. There is no significant correlation between overall self-reported arousal and performance. However, if we analyze the conditions separately, arousal in Animated condition has a significant positive correlation (Sig.=0.03<0.05) with performance, though Lecture condition does not. This can be explained by the fact that people do not tend to provide negative feedback (Charland, 2018), in this case for lecture capture. This could affect the results of self-reports and mask the link to learning.

In summary, our results suggest a significant positive relationship between emotional engagement and performance in support of H3A: The higher the emotional engagement, the better the student performance. In particular, SAM (arousal), facial expressions analysis and EDA suggest a significant relationship between mean valence / arousal and performance. The effect of variability of arousal needs further investigation as results are inconsistent between arousal measured by EDA and facial expressions. It can be assumed that both represent a different element of arousal and show different effects of those elements on performance, as discussed in literature review section (Hetland, 2016). When Harley (2015) compared facial expressions analysis results, questionnaire reports, and EDA data in a study about e-learning, he could not find a clear relationship between those measures.

5.4.2 Cognitive engagement

Our hypothesis H3B is supported and the results suggest a quadratic relationship between attention and performance (adjusted R=0.07). First, the higher the attention, the better the performance (Sig.=0.04<0.05). However, at some point, the attention has a negative effect on performance (Sig.=0.04<0.05). That effect is true regardless the condition. Table 11 provides the details. This relationship can be explained by the proposition that the visual sketchpad can be overloaded. Too much or too complex information is presented and therefore the effect of attention becomes negative on the performance as the subject is overloaded.

It is important to note that the significant relationship is only observable if we calculate the mean attention for the first ⅔ of the viewing time (500 sec). There is no significant relationship between attention and performance if we take the average of the whole video session. This can be explained by the evolution of attention over time (RQ1). Over the duration of the video, the attention significantly decreases in both conditions (β= -38 x 10 ˆ -6, Sig.=0.01<0.05). Subjects get tired, mean attention is lower at the end of the viewing period, regardless of the performance of the students. This masks the relationship between high attention and high performance. Same effect has been observed for arousal measured by EDA and discussed in Hetland's (2016) article.

Table 11: Multiple linear regression results for attention (EEG)

Variable	*β Estimate*	*t Value*	*Sig.*
Intercept	-680.62	-1.87	0.08
Mean Attention	889.34	2.09	0.04
Mean Attention (squared)	-259.85	-2.11	0.04
Condition	-2.43	-0.47	0.64

Model fit: Adjusted R2=0.07, F value=1.62, Sig.=0.21>0.05

Finally, as per regression analysis, engagement does not mediate the relationship between condition (video format) and students' performance. This outcome is expected. According to H3 results, engagement influences overall performance. At the same time, according to H2 the condition has an impact on student's performance in regard to difficult questions. There is no effect of condition on students' overall performance. Consequently, since there is no effect of condition on overall performance, there is no relationship between condition and overall performance which is mediated by engagement.

Table 12: Summary of research questions, hypotheses and results

Research questions	**Hypotheses**	**Results**
RQ1: What video production style engages more emotionally and cognitively the students?	H1A: Emotional engagement will be higher for the animated production style	Not supported
	H1B: Cognitive engagement will be higher for the animated production style	Supported
RQ2: Do we have a difference in learning between conditions?	H2: Learning performance will be higher for animated production style	Supported
RQ3: What is the relationship between student engagement and learning outcomes?	H3A: The higher the emotional engagement, the better the student performance.	Supported
	H3B: Quadratic relationship between cognitive engagement and performance	Supported

6. Discussion

Our results suggest that emotional engagement is higher for the Lecture condition when looking at the mean arousal according to EDA and facial valence evolution over time. However, the animated production style is able to better maintain the arousal of students than lecture capture. Facial arousal and arousal measured by EDA (mean and standard deviation) decrease significantly less over time in Animated condition.

As for cognitive engagement, there is a difference between the two conditions if we take time into account. The animated production style is significantly better at keeping the cognitive engagement of subjects. Also, the Animated condition shows better learning outcomes. Subjects perform overall 3% better compared to the Lecture condition. In addition, the number of correctly answered difficult questions has increased by 38% in Animated condition, and only by 18% in the Lecture condition.

Student performance is significantly correlated with self-reported arousal, facial arousal and valence, as well as arousal derived from EDA. In general, the higher the emotional engagement, the higher the performance. Also, the higher the attention, the higher the performance. However, at some point, the attention has a negative effect on performance. Thus, the relationship between cognitive engagement and performance is quadratic (an inverted U shape).

Our H1A hypothesis cannot be supported. The animated production style can maintain the arousal of subjects over time significantly better than can lecture capture. However, lecture capture video format seems to engage the subjects more emotionally at the beginning of the viewing period and students are happier to watch the lecture capture over time. This is explained by the results that show that mean arousal (EDA) is significantly higher on average but decreases significantly over time compared to the animated production style. Also, subjects are happier (valence) over time with seeing the lecture capture, but on average not significantly higher. We argue that subjects seem to be specifically aroused (mean EDA) to see the professor at the beginning of the video and respond positively (valence) over time to the social presence provided by the lecturer.

These results are supported by previous theoretical and empirical work. From the theoretical standpoint, the CTML framework suggests that the lecturer in the video will have a positive effect on students' emotional engagement by providing social cues. As for the media richness theory, the lecture capture shows the professors' body movements, which could provide additional cues, increase personal meaning (fourth criteria) and affect arousal and valence positively.

In empirical studies, it is shown that social perception is related to arousal (Lee, 2014) and that showing the lecturer increases students learning, satisfaction, and (affective) engagement (Kizilcec, 2015; Wang, 2017). In particular, when comparing lectures that include PowerPoint slides and/or no professor visuals, Kizilcec (2015) reports that students like the lecture better when the professor is shown. In our study, we see a spike in mean EDA at the beginning for the Lecture condition as the human face provides an "intimate and personal" feel compared to PowerPoint slides (Guo, 2014, p. 5).

According to our results, H1B hypothesis can be supported. The animated production style can sustain the attention and arousal of subjects to a greater extent than can lecture capture. It has interesting visual effects, animated graphics, text and audio. Thus, we can argue that the higher media richness of the animated production style allows for better emotional (arousal) and cognitive engagement over time compared to lecture capture. Previous research supports the hypothesis that richer media will provide a richer experience to learners and generate greater emotional engagement than less rich media, in our case, lecture capture (Chen, 2011). Also, based on the cinemetrics analysis, Da Silva (2016) recommends that MOOC videos should be dynamic and have short shot sequences in order to attract a student's attention.

The hypothesis of the two channels suggests that students will perform better if the multimedia learning material provides both visual and auditory stimulus compared to only one stimulus (Chen, 2015). Also, an instructor's face can provide social cues, which can evoke more attention and emotion. However, these benefits can be offset with a greater cognitive processing load, burden the students and negatively impact engagement (attention, emotion) and learning (Chen, 2011, 2015; Wang, 2017). This could especially be the case if the instructor's image does not provide additional information to that already being presented on the slides or through narration, or even takes attention away from the additional information on the slides (Homer, 2008). This does seem not to be the case for our animated production-style video, as this style is better able to preserve subject attention over time compared to lecture capture. We can argue that this type of production style better integrates visual and auditory stimuli.

The above results are coherent with the first outcome of Korving (2016) research. If the professor is visible, participants pay less attention on average during first 15 minutes than if he is not. Contrary to our study however, in Korving's study, attention paid with lecture capture decreased less over time than that for the PowerPoint-like format. According to Korving (2016), a professor's image allows students to easily pay attention and decreases cognitive processing. We assert, however, that the professor image with PowerPoint slides compared to only PowerPoint slides leads to higher attention due to split attention effects.

In summary, the animated production style provides more media richness and makes better use of visual, as well as auditory channels. It better maintains subjects' attention and emotional arousal over time. However, our results suggest that visibility of the lecturer has a positive effect on emotional engagement (arousal and valance) especially at the beginning of the lecture.

The previous findings need to be used considering that self-reported measures have their limitations (bias) (De Guinea, 2014; Sanders, 2016) or might measure different components of emotional engagement (Hetland, 2016). In a study about commercial films, Hetland (2016) compared the implicit valence measured with FaceReader and the explicit valence measured by a questionnaire. He could not find any significant relationship between those two measures. In addition, questionnaire results can be misleading due to subjects not wanting to report negative feedback and seeking to appear agreeable, as well as due to them only remembering the last part of the task (Charland, 2018). This could explain why studies that use explicit methods of measuring reported that participants were not happier watching video lectures with an instructor (Lee, 2014; Wang, 2017).

On another note, given our insignificant results on average (except mean EDA), especially pertaining to self-reported measures (SAM arousal and valence), we can understand why previous research results have not found significant differences in emotional engagement without looking at the temporal evolution of engagement.

When Chen (2015) compared different production styles (lecture capture, voice-over presentation, picture-in-picture video lectures), he found that there is no significant difference between the way these styles affect the positive and negative emotions of students. Homer (2008) stated that social presence in online learning was not significantly different with or without a professor.

As for learning, our results confirm the H2 hypothesis. Even though the animated production style allows for significantly improved performance pertaining to only difficult questions, a significant difference for all question types would have been problematic. Both videos exhibit the same content. Therefore, there should not be a significant difference between the two conditions relating to easy post-test questions. The opposite would have indicated a design issue with our videos.

This result is supported by the media richness theory. Given that the animated production style is a richer media than lecture capture, it is better at reducing uncertainty and equivocality. Thus, students are able to learn faster and better. Empirically, our results are supported by earlier research asserting that richer media improves student performance (Chen, 2011, 2015; Ilioudi, 2013). As in the case of our Animated condition, Chen (2015) outlines in his study that some formats have an improved layout, better presentation of verbal and nonverbal elements than others, and therefore are more efficient at teaching students. Also, Illioudi (2013) reported that richer media specifically allows students to master difficult topics.

According to the CTML, enhanced integration of video and audio provides for improved learning as information is processed by two separate channels (Chen, 2015). Several studies have expressed that a professor's image provides additional social cues and meaning (gesturing) and therefore increases learning (Homer, 2008; Ilioudi, 2013). When students see a professor, social interaction schemas are activated, which aids cognitive processing. At the same time, students can focus their visual attention on the lecturer and listen to audio simultaneously, which can lead to improved comprehension as information coming from two channels can complement and enhance each other (Wang, 2017).

However, our results indicate that the media richness of the animated production style allows for better learning than with lecture capture. Conversely, Wang (2017) and Homer (2008) outline in their research that the instructor image potentially does not provide significantly more information and social cues can be offset by the increased cognitive processing load of the professor's image.

Our findings confirm the H3A/B hypotheses. First, the higher the emotional engagement, the better the performance. When students are aroused, they activate their senses and resources to better attend to the content and process information. Also, decrease in anxiety and a pleasant emotional state (valance) allow subjects to focus on the learning and perform better. This result is widely supported by current literature (Chen, 2011; Harley, 2015; Homer, 2008; Lee, 2014).

As for cognitive engagement, the link between attention and learning is quadratic. Too much attention can cause a decrease in performance as the students' attentional resources get overloaded. At the same time, the first part of the inverted U-shaped relationship between cognitive engagement and achievement indicates that some students perform well but are not engaged. This could be because they are bright and were not paying attention.

These findings are in line with past literature. Korving (2016) outlines that attention is a pre-requisite for processing information and learning. Furthermore, a related study specifies that sustained attention is correlated with learning performance. Moments during the video lecture where students demonstrate low attention have a negative correlation with post-test scores (Chen, 2011). In addition, our results also align with those of Chen (2015). When too much attention is required, student performance decreases. Chen (2015) compared (a) classroom lecture, (b) voice-over presentation, (c) picture-in-picture video lectures. The voice-over presentation has the highest mean and deviation of sustained attention. At the same time, this type of video lecture has the lowest learning performance compared to the other two. Chen argues that there is a split attention effect as students have to focus on the teacher visuals, slides and table of contents at the same time. That makes it more difficult to pay attention and process information.

In summary, the video production styles have different effects on engagement. First, on average there is only a significant difference for arousal measured with EDA, in favour of lecture capture. The sustained attention is not significantly higher on average for any condition. Also, self-reported arousal and valence do not yield significant results when comparing both conditions.

When we take time into account, only facial valence increases significantly more in favour of lecture capture. Arousal measured using EDA and facial expressions, as well as sustained attention decrease significantly less over time in the case of the animated production style. Finally, as for student learning, the animated production style allows for significantly improved performance pertaining to difficult questions.

On one hand, we can argue that since the animated production style better engages the subjects emotionally (arousal) and cognitively over time, and given that there is a significant positive relationship between those states and performance, the subjects perform better in the Animated condition. On the other hand, at the beginning of the video, the subjects seem to be specifically aroused (EDA) to see the professor. However, it looks like the lecture capture can't keep the attention and arousal of subjects, even though they are happier over time with seeing the professor. Happiness does not seem to be enough to offset the decrease in arousal and attention over time and encourage additional learning. Nevertheless, the results suggest that shorter lecture capture videos could produce higher emotional and cognitive engagement, as well as improved learning than shorter animated production style videos. The best option could be to invest in rich media content while incorporating social cues.

7. Conclusion

Currently, drop out rates for MOOCs stand at around 90% (Veletsianos, 2016; Xiong, 2015). To increase retention, researchers in the field suggest choosing the most suitable video format based on how it affects the performance and engagement of students (Chen, 2015). To advance current literature and assist with that choice, our objective is to compare two currently used lecture video formats: animated production style and lecture capture. The measures of comparison are student performance and engagement.

Much of the current research on MOOC video design uses indirect metrics as a proxy for engagement of learners. However, these metrics have limitations. For instance, students don't necessarily want to finish the courses but are still engaged during instructional video viewing (Hew, 2016). There is benefit to using implicit neurophysiological measurements to understand student's engagement (Charland, 2015), as engagement evolves over time and is subject to important retrospective bias. Research is needed to investigate the progression of engagement over the duration of a video (Dillion, 2016). Also, the construct of engagement has been used very broadly in the field and a clear definition of engagement is needed for coherent interpretation. Likewise, the literature can be expanded by varying the educational context (Chen, 2011, 2015).

Our study uses Fredrick's (2004) definition of engagement. He subdivides the concept into behavioural, emotional, and cognitive engagement. First, we outline how we operationalize each of these three constructs using neurophysiological tools and measures. Then, we compare lecture capture and the animated production style by looking at engagement over time (RQ1) and student performance (RQ2). In addition, the analysis of the link between student engagement and learning outcomes is performed (RQ3). The neurophysiological variables used in this study are EEG to measure cognitive engagement, as well EDA and facial emotions analysis to quantify emotional engagement. In addition, subjects report their emotional engagement using the SAM questionnaire.

For RQ1, the two hypotheses are that emotional (H1A) and cognitive engagement (H1B) are higher for the animated production style. It should also allow for better learning outcomes (RQ2). Finally, for RQ3, our hypothesis is that the higher the emotional and cognitive engagement, the better the student performance. This relationship is quadratic in the case of cognitive engagement.

Our first H1A hypothesis is not confirmed. On average, arousal (EDA) is higher for lecture capture and over time students seem to be happier (valence) with that format. However, the animated production style is better at keeping the students aroused over time. As for the cognitive engagement, our H1B hypothesis is confirmed. The animated production style retains students' attention significantly better over time. In addition, students improve considerably more the percentage of correctly answered difficult questions in the Animated condition. Furthermore, there is a noticeable relationship between the students' performance and their emotional and cognitive engagement (H3), regardless of the condition. In general, the results suggest that the higher the engagement, the higher the performance. Too much attention, however, decreases performance at some point when the students become overwhelmed with information.

Considering these results, it seems that students perform better in the Animated condition because the animated production style is better at keeping students aroused and attentive over time. Even though subjects are happier to see the lecture capture over time and are especially aroused at the beginning of the video because they see the professors' image, these benefits are offset by the evolution of engagement over time in favour of the Animated condition. Thus, lecture capture could achieve better engagement and performance results in short videos, while the animated production style could be the better option for longer periods. A combination of the two formats (rich media content with incorporated social cues) could be the best option regardless of the length.

Future research could benefit from incorporating neuropsychological measures into study designs to gain better insight into a student's engagement and capture automatic and unconscious reactions of subjects over time. In this study, we only analyzed two different video production styles, but there are a multitude of video production styles that could be compared using our methodology. Several video production styles can be also be used in the same video, which can have various impacts compared to using only one. Additionally, our videos are 15 minutes long. It would be pertinent to see how performance and engagement vary over the duration of a complete online course.

Researchers could also perform a more granular analysis of neurophysiological data and try to understand how specific video design components affect engagement. Furthermore, while subjects in this study were mostly young students and we had a relatively small sample size, the online learning population is very diverse and large. In the future, researchers could analyse how different subpopulations engage with learning content using a larger number of subjects. Furthermore, subjects were monitored during the whole experiment, which could force engagement, and no breaks or note-taking were allowed. In real life, students follow courses online mostly at home, and subsequent studies could better mimic those conditions while capturing neurophysiological measures.

Acknowledgment and Funding

The study was financed by HEC Montreal Education Innovation fund. The funder had no decision making in research design, analysis and collection of data, or publication choice.

References

Allen, I. E., & Seaman, J. (2015). Grade Level: Tracking Online Education in the United States. *Babson Survey Research Group*.

Al-Awni, A. (2016). Mood Extraction Using Facial Features to ImproveLearning Curves of Students in E-Learning Systems. *International Journal of Advanced Computer Science and Applications*, *7*(11), 444-453.

American Encephalographic Society (1994). Guideline thirteen: Guidelines for standard electrode position nomenclature. *Journal of Clinical Neurophysiology*, *11*, 111-113.

Baddeley, A. (2003). Working memory: looking back and looking forward. *Nature reviews neuroscience*, *4*(10), 829.

Bahreini, K., Nadolski, R., & Westera, W. (2016). Data fusion for real-time multimodal emotion recognition through webcams and microphones in e-learning. *International Journal of Human–Computer Interaction*, *32*(5), 415-430.

Berka, C., Levendowski, D. J., Lumicao, M. N., Yau, A., Davis, G., Zivkovic, V. T., ... & Craven, P. L. (2007). EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks. *Aviation, space, and environmental medicine*, *78*(5), B231-B244.

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental psychiatry, 25*(1), 49-59.

Chang, T., & Chang, D. (2012). Enhancing learning experience with dynamic animation. *International Conference on Engineering Education*. University of Florida.

Charland, P. (2018). *Méthodes de recherche en neuroéducation*. Masson, Montreal.

Charland, P., Léger, P. M., Sénécal, S., Courtemanche, F., Mercier, J., Skelling, Y., & Labonté-Lemoyne, E. (2015). Assessing the multiple dimensions of engagement to characterize learning: A neurophysiological perspective. *Journal of visualized experiments: JoVE*, (101).

Chen, C. M., & Wang, H. P. (2011). Using emotion recognition technology to assess the effects of different multimedia materials on learning emotion and performance. *Library & Information Science Research*, *33*(3), 244-255.

Chen, C. M., & Wu, C. H. (2015). Effects of different video lecture types on sustained attention, emotion, cognitive load, and learning performance. *Computers & Education*, *80*, 108-121.

Courtemanche, F., Léger, P. M., Dufresne, A., Fredette, M., Labonté-LeMoyne, É., & Sénécal, S. (2018). Physiological heatmaps: a tool for visualizing users' emotional reactions. *Multimedia Tools and Applications*, *77*(9), 11547-11574.

Cronan, T. P., Léger, P. M., Robert, J., Babin, G., & Charland, P. (2012). Comparing objective measures and perceptions of cognitive learning in an ERP simulation game: a research note. *Simulation & Gaming*, *43*(4), 461-480.

Da Silva, A. G., Santos, A. M., Costa, F. A., & Viana, J. (2016). Enhancing MOOC videos: design and production strategies. *Research Track*, 107.

De Guinea, A. O., Titah, R., & Leger, P. M. (2014). Explicit and implicit antecedents of users' behavioral beliefs in information systems: A neuropsychological investigation. *Journal of Management Information Systems*, *30*(4), 179-210.

Dillon, J., Bosch, N., Chetlur, M., Wanigasekara, N., Ambrose, G. A., Sengupta, B., & D'Mello, S. K. (2016). Student Emotion, Co-occurrence, and Dropout in a MOOC Context. In *EDM* (pp. 353-357).

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of educational research*, *74*(1), 59-109.

Freeman, F. G., Mikulka, P. J., Scerbo, M. W., & Scott, L. (2004). An evaluation of an adaptive automation system using a cognitive vigilance task. *Biological psychology*, *67*(3), 283-297.

Guo, P. J., Kim, J., & Rubin, R. (2014, March). How video production affects student engagement: an empirical study of MOOC videos. In *Proceedings of the first ACM conference on Learning@ scale conference* (pp. 41-50). ACM.

Hansch, A., Hillers, L., McConachie, K., Newman, C., Schildhauer, T., & Schmidt, J. P. (2015). Video and online learning: Critical reflections and findings from the field.

Harley, J. M., Bouchet, F., Hussain, M. S., Azevedo, R., & Calvo, R. (2015). A multi-componential analysis of emotions during complex learning with an intelligent multi-agent system. *Computers in Human Behavior*, *48*, 615-625.

Hetland, A., Vittersø, J., Fagermo, K., Øvervoll, M., & Dahl, T. I. (2016). Visual excitement: analyzing the effects of three Norwegian tourism films on emotions and behavioral intentions. *Scandinavian Journal of Hospitality and Tourism*, *16*(4), 528-547.

Hew, K. F. (2016). Promoting engagement in online courses: What strategies can we learn from three highly rated MOOCS. *British Journal of Educational Technology*, *47*(2), 320-341.

Homer, B. D., Plass, J. L., & Blake, L. (2008). The effects of video on cognitive load and social presence in multimedia-learning. *Computers in Human Behavior*, *24*(3), 786-797.

Ilioudi, C., Giannakos, M. N., & Chorianopoulos, K. (2013). Investigating differences among the commonly used video lecture styles

Keltner, D., Ekman, P., Gonzaga, G. C., & Beer, J. (2000). *Facial expression of emotion* (pp. 236-49). Guilford Publications.

Kizilcec, R. F., Bailenson, J. N., & Gomez, C. J. (2015). The instructor's face in video instruction: Evidence from two large-scale field studies. *Journal of Educational Psychology*, *107*(3), 724.

Korving, H., Hernández, M., & De Groot, E. (2016). Look at me and pay attention! A study on the relation between visibility and attention in weblectures. *Computers & Education*, *94*, 151-161.

Lagerstrom, L., Johanes, P., & Ponsukcharoen, M. U. (2015, June). The myth of the six minute rule: student engagement with online videos. In *Proceedings of the American Society for Engineering Education* (pp. 14-17).

Lee, Y. H., Hsiao, C., & Ho, C. H. (2014). The effects of various multimedia instructional materials on students' learning responses and outcomes: A comparative experimental study. *Computers in Human Behavior, 40*, 119-132.

Lewinski, P., den Uyl, T. M., & Butler, C. (2014). Automated facial coding: Validation of basic emotions and FACS AUs in FaceReader. *Journal of Neuroscience, Psychology, and Economics*, 7(4), 227.

Li, X., Zhao, Q., Liu, L., Peng, H., Qi, Y., Mao, C., ... & Hu, B. (2012). Improve affective learning with EEG approach. *Computing and Informatics*, *29*(4), 557-570.

Mayer, R. E. (2005). Cognitive theory of multimedia learning. *The Cambridge handbook of multimedia learning*, *41*, 31-48.

Mikulka, P. J., Scerbo, M. W., & Freeman, F. G. (2002). Effects of a biocybernetic system on vigilance performance. *Human Factors*, *44*(4), 654-664.

Moreno, R., Mayer, R. E., Spires, H. A., & Lester, J. C. (2001). The case for social agency in computer-based teaching: Do students learn more deeply when they interact with animated pedagogical agents?. *Cognition and instruction*, *19*(2), 177-213.

Pauna, H., Léger, P. M., Sénécal, S., Fredette, M., Courtemanche, F., Chen, S. L., ... & Ménard, J. F. (2018). The psychophysiological effect of a vibro-kinetic movie experience: the case of the D-BOX movie seat. In *Information Systems and Neuroscience* (pp. 1-7). Springer, Cham.

Pope, A. T., Bogart, E. H., & Bartolome, D. S. (1995). Biocybernetic system evaluates indices of operator engagement in automated task. *Biological psychology*, *40*(1-2), 187-195.

Sanders, S. V. (2016). *Wireless EEG and self-report engagement in online learning environments* (Doctoral dissertation, Regent University).

Serrhini, M., & Dargham, A. (2017). Toward Incorporating Bio-signals in Online Education Case of Assessing Student Attention with BCI. In *Europe and MENA Cooperation Advances in Information and Communication Technologies* (pp. 135-146). Springer, Cham.

Shah, Dhawal (2016). By The Numbers: MOOCS in 2016. https://www.class-central.com/report/mooc-stats-2016/ Accessed 21 August 2018.

Sun, P. C., & Cheng, H. K. (2007). The design of instructional multimedia in e-Learning: A Media Richness Theory-based approach. *Computers & education*, *49*(3), 662-676.

Thayer, R. E. (1978). Toward a psychological theory of multidimensional activation (arousal). *Motivation and Emotion*, *2*(1), 1-34.

Trevino, L. K., Lengel, R. H., & Daft, R. L. (1987). Media symbolism, media richness, and media choice in organizations: A symbolic interactionist perspective. *Communication research*, *14*(5), 553-574.

Veletsianos, G., & Shepherdson, P. (2016). A systematic analysis and synthesis of the empirical MOOC literature published in 2013–2015. *The International Review of Research in Open and Distributed Learning*, *17*(2).

Wang, J., & Antonenko, P. D. (2017). Instructor presence in instructional video: Effects on visual attention, recall, and perceived learning. *Computers in human behavior*, *71*, 79-89.

Wang, Y. J., & Minor, M. S. (2008). Validity, reliability, and applicability of psychophysiological techniques in marketing research. *Psychology & Marketing*, *25*(2), 197-232.

Xiong, Y., Li, H., Kornhaber, M. L., Suen, H. K., Pursel, B., & Goins, D. D. (2015). Examining the relations among student motivation, engagement, and retention in a MOOC: A structural equation modeling approach. *Global Education Review*, *2*(3).

Chapitre 3 : Conclusion

L'objectif de notre étude est de comparer une vidéo de type « enregistrement de cours magistral » et une de type « enrichi ». Nous voulons savoir quel type engage le plus les étudiants et est propice à un meilleur apprentissage. En conséquence, notre première question de recherche vise à comparer les deux formats en matière d'engagement émotionnel et cognitif. Nous analysons aussi le niveau de l'apprentissage après l'expérimentation pour les deux conditions. Enfin, la relation entre l'engagement et apprentissage est examinée.

Selon les résultats des recherches précédentes, nous supposons que la vidéo de type enrichi génère un engagement émotionnel et cognitif plus élevé chez les sujets. En plus, ce type permet un meilleur apprentissage comparativement à la vidéo de type « enregistrement de cours magistral ». En effet, notre dernière hypothèse est que plus les étudiants sont émotionnellement et cognitivement engagés, meilleure est leur performance.

Pour répondre à nos questions de recherche, nous utilisons la définition d'engagement de Fredricks et précisons comment les outils neurophysiologiques et les questionnaires sont utilisés pour le mesurer. En particulier, le questionnaire Self-Assessment Manikin (SAM) (Bradley, 1994) est utilisé pour mesurer la valence et l'excitation perçues chez les sujets. En même temps, l'activité électrodermale et l'analyse des expressions faciales permettent de recueillir des données sur la valence et l'excitation implicite. Quant à lui, l'EEG donne le niveau d'attention permettant d'estimer l'engagement cognitif.

Principaux résultats

En premier lieu, les données recueillies ne soutiennent pas l'hypothèse que la capsule enrichie engage davantage émotionnellement les étudiants. La vidéo de type « enregistrement de cours magistral » est plus plaisante. Aussi, au début du visionnement, les étudiants sont plus excités lorsqu'ils regardent la vidéo. Cependant, nous pouvons constater que, dans le cas de la vidéo de type enrichi, les sujets maintiennent leur niveau d'excitation plus longtemps à travers le temps par rapport à la deuxième condition.

Selon les recherches précédentes et la théorie sur l'apprentissage avec les multimédias, nous pouvons conclure que les étudiants sont excités lorsqu'ils voient une image du professeur au début du visionnement. Grâce aux indices sociaux, ils sont aussi plus contents. Par contre, la deuxième capsule permet aux étudiants de rester plus excités à travers le temps, étant donné que le contenu est visuellement plus enrichi.

En ce qui concerne l'engagement cognitif, la vidéo de type enrichi est meilleure. Les étudiants conservent un niveau d'attention plus élevé à travers le temps. Ce résultat est en lien avec la littérature qui soutient qu'un média plus enrichi intègre mieux l'aspect visuel et cognitif pour davantage engager les sujets. Celle-là permet aussi d'apprendre plus. En analysant le questionnaire d'apprentissage, nous constatons que les questions difficiles ont été mieux apprises par les sujets qui ont regardé la capsule enrichie.

La troisième question de recherche nous permet d'intégrer les résultats précédents. Selon nos analyses, il existe un lien significatif entre l'engagement émotionnel et cognitif et l'apprentissage des étudiants. Plus les sujets sont engagés, plus la performance est élevée jusqu'à un certain niveau. Dans le cas d'engagement cognitif, trop d'engagement peut surcharger l'apprenant et diminuer la performance. En sommaire, étant donné que notre capsule enrichie engage plus les sujets à travers le temps, ils performent mieux à la fin.

Contributions à la littérature

En réalisant cette étude, nous pouvons apporter plusieurs contributions à l'état actuel de la recherche dans le domaine.

Contributions théoriques

En premier lieu, nous avons comparé une capsule de type « enrichi » avec une de type « enregistrement de cours magistral ». Ces deux types sont utilisés fréquemment pour des fins d'enseignement en ligne et n'ont pas encore été analysés pour déterminer leurs avantages/désavantages. Nos résultats contribuent à mieux comprendre ceux-ci.

En deuxième lieu, notre étude a tenu compte de l'évolution de l'engagement à travers le temps. En particulier, la moyenne de l'engagement est identique selon la majorité des mesures implicites et explicites. Par contre, si nous prenons le temps en compte, il y a des différences significatives qui nous permettent de mieux comprendre l'évolution de l'engagement à travers le temps et les différences entre nos vidéos. En troisième lieu, notre étude confirme la pertinence des outils neurophysiologiques et des variables implicites pour mesurer l'engagement. Il existe un lien entre ces mesures et l'apprentissage.

Contributions pratiques

Cette étude démontre la pertinence des méthodes statistiques d'analyse des données neurophysiologiques temporelles dans le contexte des MOOC. Ces méthodes tiennent compte du fait qu'il existe une interdépendance des valeurs recueillies à travers le temps. En revanche, la force de la relation diminue le plus lorsque les deux mesures sont séparées l'une de l'autre temporellement. Autrement dit, si une personne était engagée il y a une seconde, c'est très probable qu'elle le sera encore une seconde plus tard.

En revanche, si un sujet est engagé au début de la vidéo, ce ne sera pas nécessairement le cas 10 minutes plus tard. Étant donné que les outils neurophysiologiques génèrent des données par milliseconde, nos méthodes utilisées sont pertinentes, car ils ajustent le résultat en considérant ces dépendances. En plus, nous pouvons considérer les différences intersujets en variant l'ordonnée à l'origine de ces modèles.

Par ailleurs, nous utilisons une multitude d'outils qui captent les réactions neurophysiologiques des sujets en même temps. Cela permet de recueillir des mesures qui sont liées à plusieurs états neurophysiologiques en même temps. Ainsi, nous pouvons mieux décrire l'expérience vécue par le sujet. En plus, nous pouvons valider nos résultats en comparant les données des outils qui mesurent des états similaires, comme Biopac et FaceReader.

Limites et recherches futures

Il existe une multitude des formats de vidéos pédagogiques en ligne. Nous n'en avons analysé que deux. Aussi, nos vidéos durent environ 15 minutes, alors que les étudiants en ligne suivent des cours de plusieurs séances qui peuvent inclure plusieurs types de formats de vidéo différents. Le visionnement de ces cours s'effectue habituellement à la maison. En revanche, nos sujets étaient surveillés, donc ils ont quelque part étés obligés d'être engagés et étaient restreints dans leurs mouvements habituels. De plus, notre petit échantillon était majoritairement composé d'étudiants. En revanche, les personnes qui suivent des cours en ligne peuvent faire partie de groupes démographiques très divers.

En conséquence, les futures recherches devraient analyser d'autres types de formats de cours en ligne en tirant profit des données neurophysiologiques. Cela est envisageable lorsque ces données sont recueillies tout au long de plusieurs vidéos de formation afin de mieux comprendre la dynamique d'engagement à travers un cours complet. Les chercheurs peuvent aussi effectuer des analyses plus granulaires et essayer de comprendre comment des composants spécifiques des vidéos influencent l'engagement.

La cueillette de données pourra se déplacer du laboratoire à la maison des participants, où ceux-ci visualisent habituellement les cours en ligne tout en recueillant des données neurophysiologiques. Idéalement, ils devraient être en mesure de faire des pauses et de prendre des notes lors du visionnement de la vidéo afin de s’approcher de l’expérience réelle des étudiants. Les chercheurs pourront aussi inclure un plus grand nombre de participants qui proviennent de groupes démographiques divers pour mieux représenter les caractéristiques des étudiants qui suivent des cours en ligne.

Bibliographie

Allen, I. E., & Seaman, J. (2015). Grade Level: Tracking Online Education in the United States. *Babson Survey Research Group*.

Al-Awni, A. (2016). Mood Extraction Using Facial Features to ImproveLearning Curves of Students in E-Learning Systems. *International Journal of Advanced Computer Science and Applications*, *7*(11), 444-453.

American Encephalographic Society (1994). Guideline thirteen: Guidelines for standard electrode position nomenclature. *Journal of Clinical Neurophysiology*, *11*, 111-113.

Baddeley, A. (2003). Working memory: looking back and looking forward. *Nature reviews neuroscience*, *4*(10), 829.

Bahreini, K., Nadolski, R., & Westera, W. (2016). Data fusion for real-time multimodal emotion recognition through webcams and microphones in e-learning. *International Journal of Human–Computer Interaction*, *32*(5), 415-430.

Berka, C., Levendowski, D. J., Lumicao, M. N., Yau, A., Davis, G., Zivkovic, V. T., ... & Craven, P. L. (2007). EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks. *Aviation, space, and environmental medicine*, *78*(5), B231-B244.

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental psychiatry*, *25*(1), 49-59.

Chang, T., & Chang, D. (2012). Enhancing learning experience with dynamic animation. *International Conference on Engineering Education*. University of Florida.

Charland, P. (2018). *Méthodes de recherche en neuroéducation*. Masson, Montreal.

Charland, P., Léger, P. M., Sénécal, S., Courtemanche, F., Mercier, J., Skelling, Y., & Labonté-Lemoyne, E. (2015). Assessing the multiple dimensions of engagement to characterize learning: A neurophysiological perspective. *Journal of visualized experiments: JoVE*, (101).

Chen, C. M., & Wang, H. P. (2011). Using emotion recognition technology to assess the effects of different multimedia materials on learning emotion and performance. *Library & Information Science Research*, *33*(3), 244-255.

Chen, C. M., & Wu, C. H. (2015). Effects of different video lecture types on sustained attention, emotion, cognitive load, and learning performance. *Computers & Education*, *80*, 108-121.

Courtemanche, F., Léger, P. M., Dufresne, A., Fredette, M., Labonté-LeMoyne, É., & Sénécal, S. (2018). Physiological heatmaps: a tool for visualizing users' emotional reactions. *Multimedia Tools and Applications*, *77*(9), 11547-11574.

Cronan, T. P., Léger, P. M., Robert, J., Babin, G., & Charland, P. (2012). Comparing objective measures and perceptions of cognitive learning in an ERP simulation game: a research note. *Simulation & Gaming*, *43*(4), 461-480.

Da Silva, A. G., Santos, A. M., Costa, F. A., & Viana, J. (2016). Enhancing MOOC videos: design and production strategies. *Research Track*, 107.

De Guinea, A. O., Titah, R., & Leger, P. M. (2014). Explicit and implicit antecedents of users' behavioral beliefs in information systems: A neuropsychological investigation. *Journal of Management Information Systems*, *30*(4), 179-210.

Dillon, J., Bosch, N., Chetlur, M., Wanigasekara, N., Ambrose, G. A., Sengupta, B., & D'Mello, S. K. (2016). Student Emotion, Co-occurrence, and Dropout in a MOOC Context. In *EDM* (pp. 353-357).

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of educational research*, *74*(1), 59-109.

Freeman, F. G., Mikulka, P. J., Scerbo, M. W., & Scott, L. (2004). An evaluation of an adaptive automation system using a cognitive vigilance task. *Biological psychology*, *67*(3), 283-297.

Guo, P. J., Kim, J., & Rubin, R. (2014, March). How video production affects student engagement: an empirical study of MOOC videos. In *Proceedings of the first ACM conference on Learning@ scale conference* (pp. 41-50). ACM.

Hansch, A., Hillers, L., McConachie, K., Newman, C., Schildhauer, T., & Schmidt, J. P. (2015). Video and online learning: Critical reflections and findings from the field.

Harley, J. M., Bouchet, F., Hussain, M. S., Azevedo, R., & Calvo, R. (2015). A multi-componential analysis of emotions during complex learning with an intelligent multi-agent system. *Computers in Human Behavior*, *48*, 615-625.

Hetland, A., Vittersø, J., Fagermo, K., Øvervoll, M., & Dahl, T. I. (2016). Visual excitement: analyzing the effects of three Norwegian tourism films on emotions and behavioral intentions. *Scandinavian Journal of Hospitality and Tourism*, *16*(4), 528-547.

Hew, K. F. (2016). Promoting engagement in online courses: What strategies can we learn from three highly rated MOOCS. *British Journal of Educational Technology*, *47*(2), 320-341.

Homer, B. D., Plass, J. L., & Blake, L. (2008). The effects of video on cognitive load and social presence in multimedia-learning. *Computers in Human Behavior*, *24*(3), 786-797.

Ilioudi, C., Giannakos, M. N., & Chorianopoulos, K. (2013). Investigating differences among the commonly used video lecture styles

Keltner, D., Ekman, P., Gonzaga, G. C., & Beer, J. (2000). *Facial expression of emotion* (pp. 236-49). Guilford Publications.

Kizilcec, R. F., Bailenson, J. N., & Gomez, C. J. (2015). The instructor's face in video instruction: Evidence from two large-scale field studies. *Journal of Educational Psychology*, *107*(3), 724.

Korving, H., Hernández, M., & De Groot, E. (2016). Look at me and pay attention! A study on the relation between visibility and attention in weblectures. *Computers & Education*, *94*, 151-161.

Lagerstrom, L., Johanes, P., & Ponsukcharoen, M. U. (2015, June). The myth of the six minute rule: student engagement with online videos. In *Proceedings of the American Society for Engineering Education* (pp. 14-17).

Lee, Y. H., Hsiao, C., & Ho, C. H. (2014). The effects of various multimedia instructional materials on students' learning responses and outcomes: A comparative experimental study. *Computers in Human Behavior, 40*, 119-132.

Lewinski, P., den Uyl, T. M., & Butler, C. (2014). Automated facial coding: Validation of basic emotions and FACS AUs in FaceReader. *Journal of Neuroscience, Psychology, and Economics*, 7(4), 227.

Li, X., Zhao, Q., Liu, L., Peng, H., Qi, Y., Mao, C., ... & Hu, B. (2012). Improve affective learning with EEG approach. *Computing and Informatics*, *29*(4), 557-570.

Mayer, R. E. (2005). Cognitive theory of multimedia learning. *The Cambridge handbook of multimedia learning*, *41*, 31-48.

Mikulka, P. J., Scerbo, M. W., & Freeman, F. G. (2002). Effects of a biocybernetic system on vigilance performance. *Human Factors*, *44*(4), 654-664.

Moreno, R., Mayer, R. E., Spires, H. A., & Lester, J. C. (2001). The case for social agency in computer-based teaching: Do students learn more deeply when they interact with animated pedagogical agents?. *Cognition and instruction*, *19*(2), 177-213.

Pauna, H., Léger, P. M., Sénécal, S., Fredette, M., Courtemanche, F., Chen, S. L., ... & Ménard, J. F. (2018). The psychophysiological effect of a vibro-kinetic movie experience: the case of the D-BOX movie seat. In *Information Systems and Neuroscience* (pp. 1-7). Springer, Cham.

Pope, A. T., Bogart, E. H., & Bartolome, D. S. (1995). Biocybernetic system evaluates indices of operator engagement in automated task. *Biological psychology*, *40*(1-2), 187-195.

Sanders, S. V. (2016). *Wireless EEG and self-report engagement in online learning environments* (Doctoral dissertation, Regent University).

Serrhini, M., & Dargham, A. (2017). Toward Incorporating Bio-signals in Online Education Case of Assessing Student Attention with BCI. In *Europe and MENA Cooperation Advances in Information and Communication Technologies* (pp. 135-146). Springer, Cham.

Shah, Dhawal (2016). By The Numbers: MOOCS in 2016. https://www.class-central.com/report/mooc-stats-2016/ Accessed 21 August 2018.

Sun, P. C., & Cheng, H. K. (2007). The design of instructional multimedia in e-Learning: A Media Richness Theory-based approach. *Computers & education*, *49*(3), 662-676.

Thayer, R. E. (1978). Toward a psychological theory of multidimensional activation (arousal). *Motivation and Emotion*, *2*(1), 1-34.

Trevino, L. K., Lengel, R. H., & Daft, R. L. (1987). Media symbolism, media richness, and media choice in organizations: A symbolic interactionist perspective. *Communication research*, *14*(5), 553-574.

Veletsianos, G., & Shepherdson, P. (2016). A systematic analysis and synthesis of the empirical MOOC literature published in 2013–2015. *The International Review of Research in Open and Distributed Learning*, *17*(2).

Wang, J., & Antonenko, P. D. (2017). Instructor presence in instructional video: Effects on visual attention, recall, and perceived learning. *Computers in human behavior*, *71*, 79-89.

Wang, Y. J., & Minor, M. S. (2008). Validity, reliability, and applicability of psychophysiological techniques in marketing research. *Psychology & Marketing*, *25*(2), 197-232.

Xiong, Y., Li, H., Kornhaber, M. L., Suen, H. K., Pursel, B., & Goins, D. D. (2015). Examining the relations among student motivation, engagement, and retention in a MOOC: A structural equation modeling approach. *Global Education Review*, *2*(3).

Annexes

Annexe 1 - Detailed Data Processing

In the related work section, we have identified that results in the field are not always consistent. On one hand, Homer (2008) did not find a significant difference in social perception while comparing PowerPoint slides with and without instructor's image. On the other hand, Wang (2007) results point out the opposite. These contradictions can be caused by differences in operationalization and measurement of research variables (Kizilcec, 2015). Therefore, we need to take special care when it comes to methods and measures used.

In most studies, learning is evaluated using multiple choice questionnaires (Chen, 2011, 2015; Homer, 2008; Wang, 2017). Students' perception of different video production styles is usually measured also with questionnaires (Homer, 2008; Ilioudi, 2013; Kizilcec, 2015; Korving, 2016; Lee, 2014). Another form of measurement is log data on MOOC servers. For example, Kizilcec (2015) measures attrition by looking at the playing times of online video lectures and Guo (2014) deducts engagement of students based on their decision to proceed to assessment questions after the video lecture.

However, self-report questionnaires "do not necessarily reflect corresponding internal states of learners, even in the absence of measurement error" (Kizilcec, 2015, p. 14). This is also confirmed in other fields, such as marketing (Wang, 2008). Also, Hew (2016) outlines that server logs can be problematic in deducting student engagement. If students don't complete post-video questions, it doesn't necessarily mean that they were not engaged during the video. Several authors suggest to use innovative methodologies to gain visibility into learners internal state during video lecture viewing (Li, 2012; Sanders, 2016).

There is benefit to use implicit measures to capture automatic and unconscious reactions of subjects (De Guinea, 2014; Wang, 2008). However, only few studies capture implicit data using EEG and EDA to measure student responses during online lectures (Chen, 2015; Wang 2017). At the same time, no study until now used implicit measures to analyze the dynamics of engagement over time (the duration of the video) (Dillion, 2016). Finally, current literature suggests to use different types of measurement methods (explicit and implicit) to achieve a richer insight into learners experience (Charland, 2015; Sanders, 2016; Wang, 2008).

Data processing

We use proc mixed with autoregressive covariance structure and random-effects for the origin / intercept to analyse the neurophysiological data (over time), as outlined by Charland (2018). He specifies that random effects allow to take account for between-subjects variance in addition to intra-individual variability. This is very useful, as this type of analysis assumes that individual differences can exist in addition to experimental effects due to each condition. For example, some subjects can start with a higher cognitive engagement than others before the experiment. Random effects analysis removes those differences and produces more robust results. We fit the following models using the maximum likelihood method:

- Neurophysiological measure = $\beta\_0+ \beta\_1*$ condition+ β_2*time+ β_3*time*condition
- Time is in seconds after the start of the video
- Condition is 0 if production style is animated, and 1 if it’s lecture capture.
- Condition 0 is used as reference

Model fit is evaluated using Null Model Likelihood Ratio Test (deviance). The metric is Chi-Square.

To perform repeated-measures multiple linear regression, we need normal distribution of variables (see SAS documentation on proc mixed). To assess normality, we look at skewness and kurtosis. Variables with distributions that depart greatly from a normal distribution are normalized using log or square root transformations (Xiong, 2015) Furthermore, we account for sampling rate differences between EDA and FaceReader, similar to Harley (2015). In particular, the following computations/aggregations are performed before fitting the models.

FaceReader - Valence/Arousal: We need to aggregate our data to prevent noise which can be caused, for example, by rubbing the face (Harley, 2015). Also, “facial expression over short instants can be misleading and a time frame analysis to ascertain emotional states can provide interesting results” (Al-Awni, 2016, p. 3). Therefore, we compute an average over a 1 second for valence and arousal, similar to Hetland (2016). Missing data is then imputed by using a moving average (proc expand). No standardization of data is required, as FaceReader outputs standardized values.

EDA - Arousal: First, data is aggregated by calculating a mean over 1 second in order to remove possible artifacts (missing data, disconnections) and to have the same time scale as for FaceReader (Harley, 2015). As a second step, standard deviation over 1 sec is computed, similar to Chen’s (2015) approach. Since the variable does not have a normal distribution, we transform it (inverse, 1/x). Standardization is performed by subject (group mean centered) as suggested by Biopac and Boucsein (2011). Finally, we remove outliers as in Harley’s study (2015).

EEG – Attention: EEG raw data is extracted and imported to NeuroRT Suite (Mensia Technologies, Paris, France) for cleaning and computing the cognitive engagement ratio. Before computing the ratio, cleaning is performed according to Charland (2015, 2018) approach. Noise and artifacts, such as background electrical signal or muscular activity, need to be removed from the raw EEG data in order to measure activity related to the task.

First, we use common average reference (CAR) to reduce noise, by computing an average for all electrodes and then subtract that average from each electrode. Next, we apply a high-pass filter of 1 Hz and a low-pass filter of 100 Hz, as Mikulka (2002). With the high-pass filter we remove noise related to pulse and respiration, and with the low-pass filter we prevent aliasing. Notch filter was set at 60Hz to correct for artifacts created by electrical power lines and participants are advised to do not move more than necessary to avoid artefacts due to muscular activity (Charland, 2018). Finally, we use Mensia's blind source separation by independent component analysis (ICA) algorithm to remove a variety of artifacts including eye movements and eye blinks. After completing the cleaning step, we calculate in Mensia the engagement ratio 20 beta/(alpha + theta) according to Mikulka (2002) using F3, F4, 01, and 02 electrodes:. Each 2 seconds, the ratio is calculated over a previous 20s period (Hanning sliding window).

SAM questionnaire: First, the scale of pleasure is reversed. Consequently, the higher the value students provide on the questionnaire for any variable, the higher is the respective pleasure, arousal or control. Furthermore, we combine valence, arousal and control into a factor "SAM". Then, a mean is computed for each variable by condition. Finally, we compare the means of each variable and of the factor "SAM" using Mann-Whitney U Test as we have a small sample size, similar to Ilioudi (2013).