Master of Science in Informatics at Grenoble
Specialization: Ubiquitous and Interactive Systems

Effectiveness of video lecture design

Multimodal analysis of visual attention, affect, satisfaction, and learning outcomes

Laura Lassance de Oliveira Morais

June 26th, 2018

Research project performed at PERVASIVE Team, Inria

Under the supervision of:
Prof. James Crowley - Prof. Philippe Dessus

Defended before a jury composed of:
Prof. Laurence Nigay
Prof. Renaud Blanch
Dr. Jean-Marc Hasenfratz
Prof. François Mangenot

June 2018

Abstract

Technologies for education have substantially changed in recent decades, with increasing use of personal computers, the internet and the world wide web. The growing availability of on-line multimedia instruction, such as Massive Open Online Courses (MOOCs) mark a revolutionary new phase in this evolution. Therefore, research on the effects of multimedia instruction design on learning increasingly grows in importance. The present study conducted a pilot user experiment ($n = 24$) to evaluate which video lecture design is more effective for learning factual, conceptual, and procedural knowledge. Two video lecture designs were scrutinized: voice over slides (*without instructor* condition), and slides overlaid by *picture-in-picture* instructor video (*with instructor* condition). Among the measures, eye-gaze observational data, self-reported perceptions, and learning assessment results were analyzed. First analysis uncovered a negative effect of instructor presence on subjects' satisfaction, perceived learning, and affective response, refuting the predicted by the social-cue hypothesis. Results on learning outcomes, although not statistically significant, indicate that *without instructor* condition is more effective for learning, tending to support the interference hypothesis. Attention distribution was impacted by instructor presence: 11.6% of total fixations are over instructor in detriment of relevant data. Qualitative analysis revealed subjects look less at instructor when learning procedural knowledge if compared to factual and conceptual knowledge. In conclusion, multimedia instruction designers are provided with the guideline: "prefer voice over slides to slides overlaid by instructor starring straight at the camera and not interacting with content".

Résumé

Les technologies de l'éducation ont considérablement changé au cours des dernières décennies, avec l'utilisation croissante des ordinateurs personnels, de l'Internet et du World Wide Web. La disponibilité croissante de l'enseignement multimédia en ligne, comme les *Massive Open Online Courses* (MOOCs), marque une nouvelle phase révolutionnaire dans cette évolution. Par conséquent, la recherche sur les effets du design des supports d'enseignement multimédia sur l'apprentissage est de plus en plus importante. Dans le cadre de la présente étude, une expérience pilote a été menée auprès d'utilisateurs ($n = 24$) afin d'évaluer quel type de cours magistraux vidéo est le plus efficace pour l'apprentissage des connaissances factuelles, conceptuelles et procédurales. Deux formats de cours magistraux vidéo ont été examinés : des diapositives seules (condition *sans enseignant*) et des diapositives dans lesquelles une vidéo de l'enseignant a été incrustée (condition *avec enseignant*). Parmi les mesures, les données d'observation oculaire, les perceptions auto-déclarées et les résultats de l'évaluation de l'apprentissage ont été analysés. Une première analyse a mis en évidence un effet négatif de la présence de l'instructeur sur la satisfaction des sujets, la perception de l'apprentissage et la réponse affective, réfutant ainsi l'hypothèse de l'indece social (*social-cue hypothesis*). Les résultats à propos de l'apprentissage, bien qu'ils ne soient pas statistiquement significatifs, indiquent que la condition *sans enseignant* est plus efficace pour l'apprentissage, ce qui tend à appuyer l'hypothèse d'interférence (*interference hyopthesis*). La présence d'un enseignant a eu un impact sur la distribution de l'attention : 11,6% du total des fixations oculaires sont sur-enseignant au détriment des données pertinentes. L'analyse qualitative a révélé que les sujets regardent moins l'enseignant lors de l'apprentissage des connaissances procédurales que lors des connaissances factuelles et conceptuelles. En conclusion, nous pouvons donner aux concepteurs de cours multimédias le conseil suivant : "Préférez le format diapositives superposées par voix plutôt que d'y incruster une vidéo d'un enseignant mis en vedette par la caméra et n'interagissant pas avec le contenu".

Acknowledgements

I would like to express my gratitude to my Brazilian home university (Polytechnic School of São Paulo, USP) and the Brazilian government (Capes) for providing me with the opportunity to study in France. Also, I would like to thank Grenoble INP and UGA for accepting me and providing me with a high level of education, and enlightenment about research on Human Computer Interaction.

I am very grateful to all the prompt and inestimable assistance given by Jean-Marc Hasenfratz during the preparation of the instructional material used in this study. Jean-Marc thank you for showing to us all the MOOCs material you produced by Inria and making it available to us in no time. Also thank you very much for assisting me with the edition of the videos, it really helped me to produce a quality study.

I would also like to thank the Pervasive team and Inria for such an enriching and welcoming environment. Specially, I would like to express my profound gratitude to my two supervisors, Professor James Crowley, and Professor Philippe Dessus. James, thank you for teaching me how to build a solid research project, for the enlightening explanations about research and airplanes, for your optimism and vision, and for your weekend hours spent correcting my report. Philippe, thank you very much for the long hours devoted to explain me the bases of Cognitive Sciences, to carefully review with me every single detail of our experiment, and for always being happy to patiently answer all my questions, sometimes even more than once.

I would like to also show my gratitude to Thomas Guntz and Maxime Belgodere. Thomas, thank you for all your help with the experimental setup and the experiment itself. Thank you for always being so patient with me. And Maxime, thank you for all the times you saved me from the rain by giving me a ride, and for all the motivational and multicultural talks we had.

And last but not least, I would like to thank Kevin Richler, for all the support he gave me in this long process of research. Thank you for always believing in me, for calming me down and making me believe that everything would work fine, and for all the amazing meals that were ready at home when I arrived late from work. Without all these people this project would not have been possible.

Contents

— 1 —
Introduction

Opportunities for learning have substantially changed in recent decades. With the advent of the Internet and the world wide web, the availability of knowledge on-line and the speed with which we are presented with resources for learning mark a revolutionary new phase in education. Nowadays anyone surfing the Web can find videos tutorials on subjects ranging from how to tie a Windsor to a complete Stanford course on Einstein's Special Theory of Relativity made available to the general public in a MOOC (Massive Open Online Course) platform. The numbers illustrate this revolution. As of the end of 2017, 81 million users were enrolled in the 9.4 thousand MOOCs offered [**shah2017moocstats**]. As the scale and impact of MOOCs continues to grow, understanding how learners interact with multimedia e-learning and what is the impact of this powerful tool on learning effectiveness increasingly grows in importance. The present study is an interdisciplinary study among Educational, Cognitive and Computer Sciences, and investigates the effects of video lecture design on learning effectiveness. The final goal is to provide multimedia e-learning designers with guidelines about how to create instructional material better adapted to learners cognitive needs.

1.1 Cognitive Sciences and multimedia e-learning

Cognitive Load Theory (CLT, see, e.g., [**kalyuga1999managing**, **paas2003cognitive**, **sweller1998cognitive**]) is central to studies involving learning and problem solving. CLT explains and predicts how cognition works, that is, how information is processed by the working memory and how working memory and long term memory interact. Under CLT, three types of cognitive load (i.e., processing imposed to the working memory) are defined. *Intrinsic cognitive load* is related to the complexity of the task, in this case, the learning material. When extra information shown hinders the learning process, learners are faced with what is called *extraneous cognitive load*. Processing put into operations that allow construction of better knowledge structures is denominated *germane cognitive load*.

Mayer has built on CLT to develop a Cognitive Theory of Multimedia Learning (CTML, [**mayer2014**]), defining several multimedia learning principles. CTML is based on three assumptions: a) information is processed through two separated channels, visual and auditory; b) these channels have a limited capacity of processing; and c) learning is an active process of filtering, selecting, organizing and integrating information. Mayer's principles shed light on how to design multimedia instructions that maximize the integration of new information to prior knowledge, i.e., of successful learning. For example, the image principle states that "people do not necessarily learn better from a multimedia lesson when the speaker's image is added to the screen."

However, empirical findings in the domain of multimedia learning are not unanimous about which choice of video lecture elements are effective for learning. Therefore, due to the ubiquitous presence of multimedia instruction in the current education scenario, it is increasingly important to have a research effort towards better defining and understanding the effects of multimedia e-learning design on learning experience and outcomes.

1.2 Effectiveness of video lecture design

The present study focus on answering the question:

Research Question: Which is the most effective video lecture design for learning different types of knowledge?

Effectiveness for learning refers to both learners satisfaction with the learning experience, and their learning outcomes. MOOCs have the distinguishing characteristic of being non-obligatory and of having their pace completely controlled by the learners. That means that learners will only follow the entire MOOC if they are motivated by the course and satisfied with their learning experience, making essential the inclusion of learners' satisfaction as a measure of learning effectiveness. On the other hand, the final goal of any instruction, lecture or course is that people learn the content being taught. Therefore, it is also indispensable to measure learning effectiveness through learning outcomes.

Different types of knowledge may be most effectively expressed with different forms of multimedia presentation. Thus, any guidelines for multimedia learning must consider the types of knowledge that are to be learned. Three types of knowledge were taken into account in this study: factual, conceptual, and procedural learning. In terms of video lecture designs, a broad variety was considered (*cf.* [**crook2017video**]). However, due to the availability of sample material, only two of these were included in this study: voice-over slides (*without instructor* condition), and picture-in-picture instructor video overlaying slides (*with instructor* condition).

Two theories about the effects of instructor presence on learning are currently accepted and studied in the field. The social-cue hypothesis defends that the nonverbal communication cues provided by the video of the instructor evoke the social interaction schema [**clark2016learning**] on learners, leading learners to engage in deeper cognitive processing and consequently allowing better learning outcomes. Under the social cue hypothesis, the social agency theory [**mayer2003social**, **moreno2001case**] also enounces positive effects of instructor presence on social-affective response. On the other hand, the interference hypothesis considers there are two conflicting sets of visual stimuli: the instructor video, and the rest of the screen showing the support material with text, graphs, and images. This predicts that the instructor presence may overload the processing capacity of the visual sketchpad, thus the visible instructor may generate extraneous cognitive load, hindering learning.

In view of these two conflicting theories, three groups of experimental hypotheses were investigated (for detailed formulation: section 2.3). Under the social-cue hypothesis, it is firstly postulated that instructor presence will have a positive effect on learners by stimulating engagement, satisfaction, willingness to continue following the MOOC, affective response, and perceived learning. We also investigated this in a comparative study between groups of learners from different educational profiles (i.e., Informatics, and non-Informatics). Our hypothesis was that non-Informatics learners would have better responses to instructor presence than Informatics learners, due to the specificities of the field [**santos2016speakers**]. A second research question was posed concerning learning outcomes. In view of inconclusive results in the published literature, we sought to determine which of the two conflicting theories is validated for learning outcomes: the social-cue hypothesis or the interference hypothesis. In addition we analyzed the attention distribution of learners, under the hypothesis that the presence of the instructor takes an important amount of learners' attention during the video lecture. To address all the hypotheses, a pilot experiment was conducted, containing observational, self-reported, and learning outcome measures necessary to the study of the hypotheses.

In organizing this study we have chosen a broader and somewhat unexplored approach. We investigate both learner satisfaction and the effectiveness of video lecture designs as a function of the type of knowledge that is taught. As is common in this field (e.g., [**kizilcec2014showing**]), we have used eye-tracking as an observational measure of subjects' attention dynamics, in an attempt to understand the effects of design elements.

1.3 Experimental measures, protocol, and results

An experiment protocol was created to evaluate effectiveness of video lecture design on learning different types of knowledge. Subjects first answered a profiling pre-questionnaire, than watched a 10 minutes video lecture about Web of Linked Data and Semantic Web, next subjects answered an affective post-questionnaire and solved questions from a learning assessment test. In the end, they were asked for their thoughts and impressions in an open conversation, and given an illustrative application example of the course content. For the viewing session, participants were given 15 minutes to watch the video (i.e., 150% the lecture duration), extra time they could profit to use the learners controls available to them (pause, next chapter, previous chapter, and change position in time bar). There were two video lecture designs available, i.e., two experimental conditions: voice over slides (*without instructor* condition), and instructor video overlaying slides, either in a mobile frame or shared presence on the screen (*with instructor* condition). Subjects were classified into two educational profiles: Informatics, and non-Informatics. It was considered that people having studied or working on the same field as the one from the video lecture content might respond differently to the lectures, imposing educational profile as a controlled parameter. The experiment was carried out with 24 voluntary participants, and had balanced independent groups ($n = 6$) for each of the four combinations of educational profile and experimental condition.

In order to record observational data, the experiment counted with a setup composed by a touch-screen 23.8" computer, a Kinect 2.0, a frontal 1080p Webcam, and a Fovio Eye-Tracking bar. On-ear headphones were used to provide a more intense feeling of immersion in the task. The experimental setup provides data about subjects eye gaze, body posture, and affect, allowing a study of subjects behavior while watching the lecture. In addition, the affective post-questionnaire collected information about subjects' self-reported engagement, satisfaction, willingness to continue following the MOOC, perceived learning, affective response, and mental effort put into watching the lecture. These measures are based in experiments of other studies in the field [**wang2017instructor**, **kizilcec2014showing**, **paas1993efficiency**]. Self-reported mental effort put into solving the questions is also asked from participants after they finish the test. The learning assessment test produces an objective measure of learners' learning process outcomes. Three levels of learning outcomes are tested, from the simplest to the more robust, respectively: recognition, retention, and transfer. Furthermore, in order to distinguish viewing behavior and learning per types of knowledge, the video lecture is divided into 16 chapters that are classified into factual, conceptual, or procedural knowledge. The three types of knowledge are balanced in the learning assessment questions for all three levels, and reasonably in the video lecture in terms of topics covered (factual knowledge: 3 main and 4 secondary topics, 7 in total; conceptual knowledge: 4 main topics; procedural knowledge: 3 main and 2 secondary topics, 5 in total).

In the analysis phase, the learning assessments were corrected, and their points normalized, attributing the same weight for each content topic tested. For opened questions, the corrections were validated for two people. The learning scores are classified into four types: recognition [0-1], retention [0-1], transfer [0-1], and total [0-3], being the latter the sum of the three first. The three learning outcomes levels have the same weight in a simplification effort, and also as their are three equally necessary parts of the learning process. A total of 912 data points were collected from the questionnaires and the test, considering all participants. In terms of observational data, around 35 Gb of data was recorded and stored for each participant.

Statistical analysis was performed over the data in order to evaluate the hypotheses. The statistical tests employed were the comparison of means through unpaired two-sampled t-tests (one- or two-tailed, depending on the hypothesis), factorial ANOVAs to compare influence of the two control variables (i.e. experimental condition and educational profile) on the mean differences of a measure, and Pearson's correlation for analysis or relationship between numerical variables, such as perceived learning and learning outcomes. The significance level is set to $\alpha = 0.05$, and for $0.05 < \textit{p-value} < 0.10$ a tendency is considered detected. The normality and homogeneity of variances assumptions were verified before every test, and, in case they were not verified, a correspondent test was carried out (difference of means: unpaired two-samples Wilcoxon test; correlation: Kendall's tau coefficient). In addition, the analysis of attention distribution was made using the defined Areas of Interest (AOIs) in the condition *with instructor*. In each slide an AOI was defined for the instructor, and others for the content. Heat maps of subjects' fixations were generated and gave place to a qualitative analysis of attention distribution for different types of knowledge.

The results obtained for learners' satisfaction, perceived learning, and affective response refute the social-cue hypothesis, that is, instructor presence had negative effects on some of learners' perceptions. The learning outcome results, although not statistically significant, tend to support the interference hypothesis, that is, the instructor promoted extraneous cognitive load and hindered learning. Therefore, the present study tend to support the interference hypothesis. In addition, a fair impact of instructor presence on learners' attention distribution was detected, with an average of 11.59% of the total fixations being dedicated to the instructor AOI. The qualitative analysis of fixation heat maps suggested that attention is given to instructor in detriment of the content, and that, for procedural knowledge, less attention is given to the instructor's AOI. Therefore, this first analysis of the data collected in a pilot study suggests the following guideline for multimedia instructor designers:

Guideline: Prefer voice over slides video lecture design to slides overlaid by a *picute-in-picture* video of an instructor starring directly at the camera and not interacting with the content.

The present report marks the end of the first stage of this project. A second stage will be developed by the student Laura Lassance in her home University in Brazil, as her Brazilian's bachelor final project (4 months). There, the data about pupil dilation, facial expression, and body postures will be analyzed and crossed against the self-reported measures. The research will aim at answering if the experimental setup used in this study can be used to accurately determine viewers' cognitive load while learning. Accuracy will be determined by equivalency and redundancy of observational data, for example, cognitive load determined by pupil dilation crossed against the one detected by facial expressions, and also equivalency to self-reported cognitive load measures present in the dataset.

1.4 Contents of this report

Chapter 2 presents a review of the state of the art in multimedia e-learning research, and defines the research question and experimental hypothesis explored in this study. We seek to answer which video lecture design is more effective for learning different types of knowledge. This section details the two taxonomies used as basis for the experimental design choices. First, the three types of knowledge considered, i.e., factual, conceptual, and procedural knowledge are detailed. Then, Crook et al.'s taxonomy of video lecture designs is presented, revealing which design can currently be found in the multimedia e-learning scenario. Regarding the theoretical framework behind this research, Cognitive Load Theory (CLT) introduces how working memory processes information and how it interacts with long term memory. The definition of intrinsic cognitive load, extraneous cognitive load, and germane cognitive load are stated, and their relation to instructional design choices is explained. Mayer's Cognitive Theory of Multimedia Learning (CTML), built on CLT, is briefly stated, exposing how auditory and visual channels process information, and the limited capacity of working memory. Multimedia learning empirical evidence is inconclusive about what is the effect of instructor presence on learning experience and outcomes. Some published literature support that instructor presence has a positive effect on both learners' affective response to the material and learning performance. These findings support the effects predicted by the social agency theory, validating the so called *social-cue hypothesis*. Other studies uncovered evidences that having the instructor visible, mostly if he is starring straight at the camera, impact negatively learners, producing negative affective response, and hindering learning. These results are framed in the *interference hypothesis* that agree state the detrimental overload of the visual sketchpad causing the split attention effect, and consequent negative impact of instructor presence on learning. Our experimental hypothesis are posed, addressing three aspects of multimedia e-learning: learners' engagement and perceptions about the learning experience, effectiveness in terms of learning outcomes, and impact of instructor presence on learners' viewing behavior and attention distribution.

In chapter 3, the pilot user experiment conducted is presented, detailing the setup used, the experimental protocol, and which measures were selected to evaluate the experimental hypotheses posed in the present study. Twenty four voluntaries participated in the experiment. Twelve have an Informatics educational profile, and the other twelve are from other fields, and, thus, regrouped under the non-Informatics profile. The experiment counted with a profiling pre-questionnaire, a 10 to 15 minutes viewing session of the video lecture selected, a post-questionnaires about learners' perceptions of learning experience, and a learning assessment test. The video lecture used is a 10 minutes Web of Linked Data video lecture in French offered in two designs: voice-over slides, and visible instructor overlaying slides. During the 5 extra minutes subjects could make use of the learners controls available, i.e., pause, next and previous chapter, and manipulation of time bar. The experimental setup recording the viewing session is composed of a 23.8" touch screen computer, a Kinect 2.0, a frontal 1080p Webcam, and a Fovio bar eye-tracker. This equipment records observational data, allowing objective measures of attention distribution (eye-gaze), affective response (facial expressions), and learners' behavior (body posture, and facial expressions). Subjective self-reported measures were also collected through the application of questionnaires and recognition, retention, and transfer were objectively measured with a learning assessment test. Chapter 3 also presents the 16 chapters of the video lecture classified into either factual, conceptual, or procedural knowledge. The learning test covered a balanced amount of factual, conceptual, and procedural knowledge topics. In the end of the chapter, the protocol used during experiments is detailed step by step.

Chapter 4 shows how the 912 data points from the questionnaires and the approximately 837 Gb of observational data were treated and analyzed in order to evaluate the experimental hypotheses. The grading process and the normalization of recognition, retention, and transfer scores are detailed. There are four independent and balanced groups of subjects ($n = 6$) formed by the combination of the two control variables, i.e., experimental condition and educational profile. Statistical and qualitative analysis was performed over the data. Statistical tests such as t-test and Pearson's correlation were applied to compare the effects of the controlled variables over the measurements. The significance level was set at $\alpha = 0.05$, and in cases where $0.05 <$ *p-value* < 0.10 a tendency was considered to have been found. An example is the comparison between the means of subjects' satisfaction scores across the two experimental conditions. The significancy test returned that the satisfaction score mean of *without instructor* condition was greater than the mean for *with instructor* condition with a *p-value* of, approximately 0.060 ($<$ 0.10). Hence, we concluded that instructor presence tend to have a negative effect on subjects' satisfaction. The evaluation of each experimental hypothesis is demonstrated in chapter 4, with detailed information about the measures and statistical tests used. In addition, for analysis of instructor presence on subjects' attention distribution, areas of interest (AOIs) were delineated around the visible instructor and around relevant content. The portion of fixations detected in instructor AOI was combined to heat maps of subjects' eye fixations in order to carry out a qualitative analysis of attention distribution per type of knowledge.

At last, chapter 5 discusses the obtained results, comments on the study limitations, and introduces the research outlook. An initial analysis uncovered a negative effect of instructor presence on subjects' satisfaction, perceived learning, and affective response, refuting the social-cue hypothesis. In terms of learning outcomes, although statistically non significant, results indicate that for recognition the instructor presence may be beneficial, however for retention and transfer subjects in the *without instructor* condition achieved better performances. Therefore, initial guidelines for instructional designers are provided based on the tendency of this study to support the interference hypothesis. In addition, we found results similar to the literature about learners' attention distribution when instructor is visible. Instructor AOI received, in average, 11.59% of subjects' fixations in detriment of relevant knowledge. It was also uncovered that when learning procedural knowledge subjects look less at the instructor than when watching factual and conceptual knowledge chapters. Chapter 5 also presents limitations of the experiment, such as number of participants, and variety of video lecture designs tested. This research only evaluated the presence of a instructor starring straight at the camera and not interacting with content, what might not provide enough nonverbal communication cues to observe the state on the social-cue hypothesis. Simple modifications to the experimental protocol are also suggested, such as rephrasing the perceived learning question due to misinterpretation, and measuring also the time subjects take to solve the questions as a measure of cognitive load. The conclusion chapter finishes suggesting further analysis of the observational data in order to determine if the experimental setup used in this study could be used to measure viewers' cognitive load in an on-line manner. In case affirmative, it could be used to create an adaptive learning environment. This exploration is object of the second stage of this project, to be developed by the student Laura Lassance in her Brazilian home University.

— 2 —

Evaluating the design of video instruction

Technologies for education have substantially changed in recent decades, with increasing use of personal computers, the internet and the world wide web. The growing availability of online course material marks a revolutionary new phase in this evolution. In 2012, during its first year of operation, Coursera, registered 2.9 million users from more than 220 countries enrolled in 328 courses [**waldrop2013campus**]. Over the last few years, the availability and popularity of online course material have continued to enjoy rapid growth. According to data collected by Class Central[1], a search engine and reviews site for MOOCs, 81 million participants have enrolled in MOOCs as of the end of 2017 with around 9.4 thousand courses offered [**shah2017moocstats**].

As the scale and impact MOOCs continues to grow, the Educational Sciences research community has increasingly turned its attention to understanding this new tool. The MOOC format makes it possible to complement video lectures with extra course material. Engagement is driven and stimulated by social tools such as discussion forums, virtual office hours, and prompt instructor feedback. The result is a learning environment that greatly enhances social interaction approaching traditional classroom learning environments. For example, in discussion forums students present themselves, share experiences, as well as promote discussions about lecture topics, assignments, and supplementary course materials.

A variety of research has studied the experience of learners in MOOCs. Many investigators have concentrated on how social interaction stimulates learning while others have explored the effectiveness of structure and pace in online courses. Video lectures play a central role in MOOCs, as shown by Grünewald [**grunewald2013designing**] (depicted in Figure 2.1). Video course material are found in a variety of layouts (Figure 2.3), imposed by the search of which lecture design is most effective for maximizing learning and engagement [**kizilcec2014showing**]. Research has been done to better understand the effects of lecture design elements on students' learning experience and outcomes.

Many researchers have focused on presence of an instructor. Two prevalent theories in this area offer contradictory guidance. According to the Cognitive Theory of Multimedia Learning (CMTL, [**mayer2014**]), social agency theory [**mayer2003social**, **moreno2001case**, **cui2013building**], and the activation of a social interaction schema [**clark2016learning**] predict that the presence of an instructor has a positive effect on learning experience. On the other hand, Cognitive Load Theory (CLT, see, e.g., [**kalyuga1999managing**, **paas2003cognitive**, **sweller1998cognitive**]) and the split attention effect [**antonenko2010influence**, **kalyuga1999managing**] would seem to indicate that the presence of an instructor has a negative impact on the effectiveness of learning. The present study was motivated as an attempt to resolve these conflicting predictions, and to explore the problem of effectiveness of video lecture designs

In organizing this study we have chosen a broader and somewhat unexplored approach. We investigate both learner satisfaction and the effectiveness of video lecture designs as a function of the type of knowledge that is taught. As is common in this field (e.g., [**kizilcec2014showing**]),

[1] Class Central: `https://www.class-central.com/`

we have used eye-tracking as an observational measure of subjects' attention dynamics, in an attempt to understand the effects of design elements.

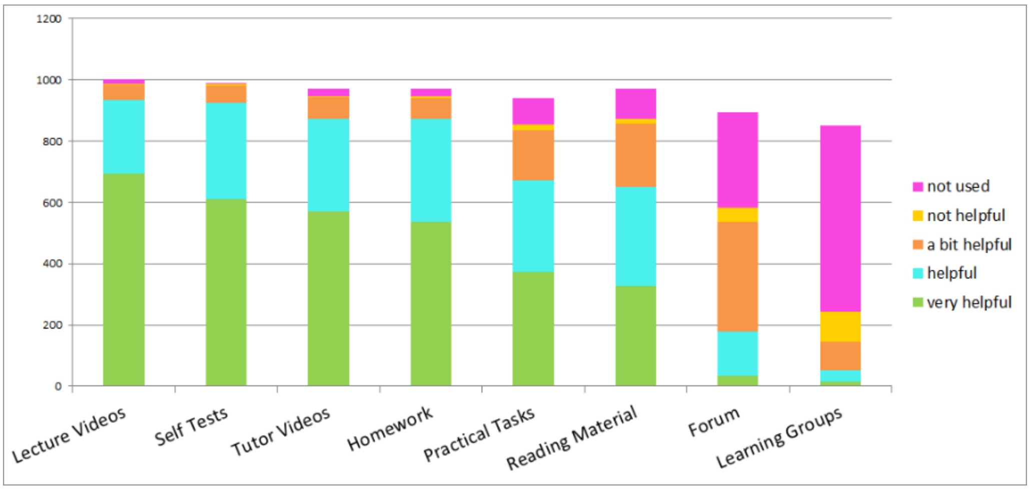


Figure 2.1 – Judgement of usefulness of the different types of learning materials present in a MOOC. This study was carried out with people (n=1,153) enrolled in the "Internet working with TCP/IP" MOOC offered in 2012, in German by openHPI, the MOOC platform of the Hasso Plattner Institute (HPI), in Potsdam , Germany.

2.1 Taxonomies of types of knowledge and video lecture designs

Two taxonomies were used as basis for the present study. One, regarding the types of knowledge, is presented by Anderson et al [**anderson2001taxonomy**]. The other is Crook and Schofield's taxonomy of online lecture design [**crook2017video**].

Types of knowledge

Types of knowledge have different functions in the performance of a task [**de1996types**], so it is reseasonable to ask that instructional design differ for different types of knowledge [**merrill2000knowledge**]. To address this question, we chose to control two parameters: type of knowledge, and video instruction design. This way we are able to reach answers more adaptable to different domains.

We chose a revision of Bloom's taxonomy [**anderson2001taxonomy**] as our classification system for the knowledge types. The taxonomy divides the knowledge dimension into 4 major types: *Factual*, *Conceptual*, *Procedural*, and *Metacognitive*.

Factual knowledge is the knowledge of isolated pieces of information, such as terminologies, dates, names of places, and descriptive details about facts. In contrast, *conceptual knowledge* comprehends more complex and structured information. Concepts are usually formed by interconnections among chunks of knowledge. As an example, consider knowledge about birds. Knowing that birds fly, and that they sing to attract a mate are examples of factual knowl-

Major Types and Subtypes	Examples
A. Factual knowledge -	The basic elements students must know to be acquainted with a discipline or solve problems in it
Aa. Knowledge of terminology	Technical vocabulary, musical symbols
Ab. Knowledge of specific details and elements	Major natural resources, reliable sources of information
B. Conceptual knowledge -	The interrelationship among the basic elements within a larger structure that enable them to function together
Ba. Knowledge of classifications and categories	Periods of geological time, forms of business ownership
Bb. Knowledge of principles and generalizations	Pythagorean theorem, law of supply and demand
Bc. Knowledge of theories, models, and structures	Theory of evolution, structure of Congress
C. Procedural knowledge -	How to do something, methods of inquiry, and criteria for using skills, algorithms, techniques, and methods
Ca. Knowledge of subject-specific skills and algorithms	Skills used in painting with watercolors, whole-number division algorithm
Cb. Knowledge of subject-specific techniques and methods	Interviewing techniques, scientific method
Cc. Knowledge of criteria for determining when to use appropriate procedures	Criteria used to determine when to apply a procedure involving Newton's second law, criteria used to judge the feasibility of using a particular method to estimate business costs

Table 2.1 – Taxonomy of knowledge types. Classification extracted from the revision of Bloom's taxonomy [**anderson2001taxonomy**], page 29.

edge. However, seeing something flying or hearing it singing and being able to recognize it is a bird is part of understanding a piece of conceptual knowledge.

There is also the knowledge of the "how to", that is, the *procedural knowledge*. It not only includes the ability to use algorithms, techniques, and methods, but also the understanding of how to determine when to use them. Finally, *metacognitive knowledge* is the understanding about cognition in general as well as awareness and understanding of one own's cognition. This involves being able to judge, for example, being able to reflect on one's own learning process and identify why one did not learn a certain content matter.

Three types of knowledge were considered in our study: *Factual*, *Conceptual*, and *Procedural*. Table 2.1 summarizes the selected knowledge types and their associated subtypes. Of these, metacognitive knowledge was not included in our experiment, as this form of knowledge is much less widely used in online courseware. In addition, the formulation of questions to assess metacognitive knowledge is highly complex, and assumes a well developed understanding of cognition by the learner.

Taxonomy of video lecture designs

A variety of different layouts may be used in constructing an instructional video. Classifications for the designs of video lectures have been created as means of organizing research on the efficacy of multimedia e-learning. Crook and Schofield classified 16 formats of video design into 5 main categories [**crook2017video**]. Santos-Espino et al. identified two tendencies in MOOCs' video styles [**santos2016speakers**]: *speaker-centric*, and *board-centric*. The usage of these approaches differ for different course subject's areas. STEM fields (Science, Technology, Engineering, and Mathematics) use more the board-centric approach, whilst Arts and Humanities apply mostly speaker-centric styles.

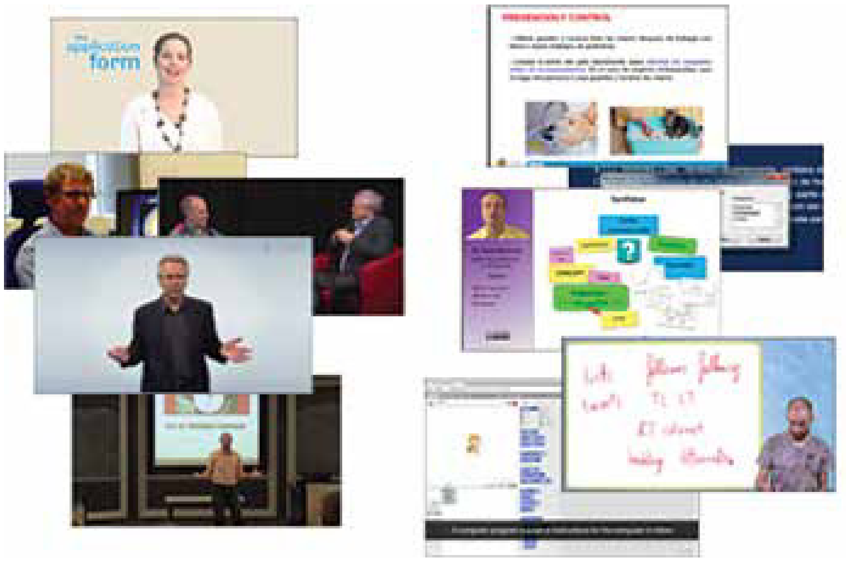


Figure 2.2 – Speaker-centric designs (left) and board-centric designs (right) [**santos2016speakers**]

Crook and Schofield's taxonomy of video lecture designs is used as the stepping stone for the video lecture designs considered in the study. In order to classify the sixteen formats identified, two elements were analyzed. The narrator, that is, the principle voice or instructor is the first element. The second is the visual context in which the narrator is put on: "domestic" scene, series of slides, whiteboard, or topic-relevant context. Figure 2.3 shows the resulting taxonomy of video lecture designs (full categories definitions: refer to Appendix F). The main categories are coded with letters from A to E. Theirs subdivisions are identified by numbers.

Across the taxonomy, a rough continuum of designs regarding the balance between the lecturer presence and the content can be detected. Here "content" means slides, whiteboard, or any graphical representation of the content. Row A starts with content-heavy and lecturer-light formats. The balance is reversed in row E, where there is only the lecturer(s) delivering the content with no graphical aid. There is also a progression of narrator interaction with the content. Such interaction has a positive impact in learning. As it directs learners' attention to the most import and relevant aspects of the content (signaling function [**mayer2014**, **van201411**]), such nonverbal cues lead to better recall performance [**wang2017instructor**].

Santos-Espino et al. identified 7 video styles as the most frequent in MOOCs ($N = 115$) [**santos2016speakers**]: *Talking Head*, *Live Lecture*, *Interview*, *Slides*, *Screencast*, *Virtual White-*

board, and *Documentary*. We based our subjects' educational background profiles on the subjects areas considered in this study: arts and humanities (*Hum*), business and management (*BMg*), social sciences (*Soc*), Health and Medicine (*HMed*), natural sciences and mathematics (*Sci*), engineering and technology (*Tech*), and other (*Other*).

Board-centric (mostly Slides) or mixed (mostly between Slides and Talking Head, that is, *Head and Slides*) styles prevail in STEM courses. These style preferences may be related to characteristics of the content thought. For example, STEM courses often rely on mathematical proofs and demonstrations, that can be better explained using boards. In contrast, humanities courses rely more on narratives, which can appropriately be heard without continuous and simultaneous text or image display. Hence, as we focus in Sci and Tech fields, and the styles *Slides* and *Head and Slides* are considered for the video lecture to be used in the experiment. Finer video lecture design filtering was based on Crook and Schofield's taxonomy.

2.2 Theoretical framework and empirical evidence

A video of the instructor is included in various instructional videos as a way to simulate a social interaction with the teacher. Instructor deictics may also serve as signaling and direct learners' attention to the relevant content. The design decision of having a visible narrator attempts to increase learner motivation and engagement in online learning environments. However, theoretical base and empirical evidence for the support of such design choice are limited and mixed. The image principle of the Cognitive Theory of Multimedia Learning illustrates well this scenario. It states that presenting the speaker's image not necessarily leads people to deeper learning [**mayer2012embodiment**, **mayer2003multimedia**]. However, there is evidence that reported engagement is enhanced in condition with talking head of the instructor, when compared to condition without it [**guo2014video**, **kizilcec2014showing**]. Therefore, the theories and mixed empirical evidences support both positive and negative effects of instructor presence in video lectures. Both are discussed in this section.

Learning and engagement: the social-cue hypothesis

The Cognitive Theory of Multimedia Learning (CTML, [**mayer2014**]) is the foundation for research about effectiveness of video lecture design features. CTML is based on the model that human comprehension relies on at least three forms of memory: sensory, working, and long-term memory. The learning process engages all three forms. First, one attends to verbal explanation and visual representation of a given task through ears and eyes. Then, verbal and visual information is actively selected (attention) from the sensory memory, and enter the working memory. This information is organized in the working memory into coherent mental representations, that is, schemas. Within working memory, both verbal and visual schemas will be integrated between themselves and with prior knowledge retrieved from long-term memory. These mental representations are encoded into the long-term memory. If meaningful learning took place, this information can be retrieved as an unique chunk once one needs to perform a given task.

Integrating knowledge into a sole schema is important, because working memory capacity is limited [**baddeley1974working**], imposing a bottleneck for learning. Working memory can hold seven plus or minus two chunks of information [**miller1956magical**], and only four items

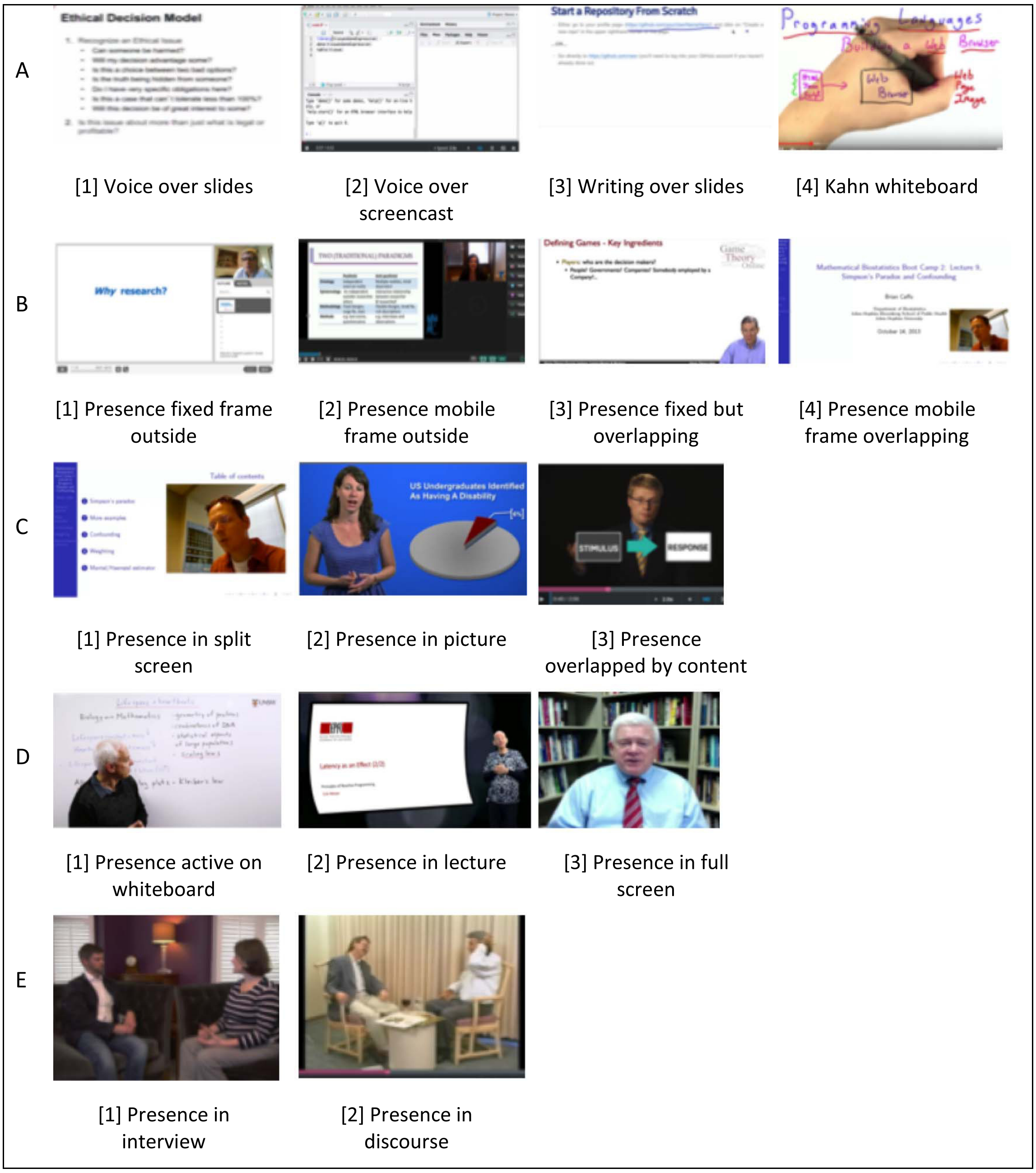


Figure 2.3 – Taxonomy of video lecture designs (taken from [**crook2017video**], page 60). Two elements analyzed: narrator and visual context of its presence. The narrator is either the principle voice, or the lecturer. The narrator's visual context may be a "domestic" scene, a series of slides, a whiteboard, or a topic-relevant context. For categories detailed definitions refer to Appendix F.

at a time processing information [**cowan_2001**]. Working memory model ([**baddeley1974working**]) states that there are different units for different types of information. Central executive unit manages the processing and allocates the information to the correct specialized unit. Verbal stimuli is treated in the phonological loop. Visual stimuli is stored and processed by the visuospatial sketchpad. In multimedia e-learning context, the professor narration is the phonological stimuli, while information displayed on screen is treated by the visuospatial sketchpad.

A complex visual stimuli present in many video lectures is the video of the professor. It provides nonverbal communication cues, such as mutual gaze, gesturing, and facial expression, known to play an important role in interpersonal interaction [**argyle1988bodily**]. According to the social agency theory [**mayer2003social**, **moreno2001case**], social cues activate the social interaction schema, causing learners to engage in deeper cognitive processing. Hence, the presence of video or image of the instructor may result in learners processing lecture content in a deeper level [**clark2016learning**]. Furthermore, instructor presence in multimedia learner may also have a socio-emotional positive effect. Such beneficial responses, under social agency theory, are induced by learners' feeling of being in an interaction with another person due to the verbal and nonverbal social cues provide by instructor presence in the video lecture [**cui2013building**].

Several studies sought empirical evidence for the social-cue hypothesis, that states that adding instructor presence to multimedia instruction has positive effects on learners' perceptions and learning performance. However, their results are inconclusive, imposing the need for other approaches to research the topic. For example, Chen and Wu [**chen2015effects**] identified superior learning performance for subjects in conditions where the instructor is visible (lecture capture, or picture-in-picture) when compared to the voice-over format. However, no significant emotional response differences were detected across the three video instruction formats (lecture capture, picture-in-picture, and voice-over). Homer and colleagues [**homer2008effects**] found no significant effects of instructor presence, neither on learning outcomes nor on social presence perception. Other studies found that instructor presence had positive effects on learners perceptions, but no difference in learning outcomes was detected. One of them, conducted by Kizilcec et al. [**kizilcec2014showing**], investigated the effects of adding a small video of the instructor over the slides on learning experience and performance. They uncovered only an affective benefit of such design element, that was thought to have educational value, but its inclusion did not impact recall ability (short- and medium-term). Another one, by Wang and Antonenko [**wang2017instructor**], also explored the impact of instructor presence on learners' perceptions, and learning outcomes (recall and transfer), but now for different difficulty levels: an easy and a difficult topic. For the difficult topic, results were similar to Kizilcec et al. ones, that is, perceived learning and satisfaction were significantly higher when instructor was present, but recall ability was similar for both conditions. However, easy topic results were closer to social-cue hypothesis, because both perceptions (perceived learning, satisfaction) and learning (recall) were enhanced by the presence of the instructor. Nevertheless, for both difficulty levels, transfer ability neither improved nor was hindered by picture-in-picture instructor presence. This illustrates well the statement that results for social-cue hypothesis are conflicting.

Besides the use of experimental design, research to understand influence of instructor presence was also conducted in ecological conditions (i.e., in a non-controlled environment, the "real-life" situation). Guo et al. [**guo2014video**] analyzed 6.9 million video watching session across four courses attended on edX MOOC platform. Video watching sessions were examined for engagement (i.e., video watching session length and attempts at post-video assessment problems). Conclusions were that when video instruction with voice-over was interspersed by talking head of instructor learner engaged more than in condition with only narration over instruction material. In addition, a study based on clickstream data of one Coursera course detected engagement enhancement when instructor seamlessly interacted with content shown in the video, when compared with instructor present in a fixed window condition [**bhat2015seeing**]. Preference for and better learning experience with interaction condition was attributed to the more realistic cues offered by such seamless interaction with the content.

Cognitive load and split attention: the interference hypothesis

The main postulate of Cognitive Load Theory (CLT) is that instruction effectiveness arises from designing instruction in alignment with learner's cognitive architecture [**van2008instructional**]. Briefly stated, CLT explains and predicts how cognition operates, that is, how information is processed in working memory and how these processes interacts with long-term memory. Cognitive load is defined as the load imposed on working memory to manipulate and store information being presented ([**sweller1998cognitive**]; for more details on CLT see also, e.g., [**sweller1988cognitive**, **van2005cognitive**]). Cognitive Load Theory distinguishes three types of cognitive load:

- *Intrinsic cognitive load* is related to the complexity of the instructional material, that is, the number of interacting information elements presented. For learning to start working memory needs to processes all interacting elements simultaneously [**chandler1991cognitive**]. Hence, intrinsic cognitive load is moderated by learners' prior knowledge, because information priorly known was already organized into schemas and are brought to processing as a single chunk in working memory. Therefore, the number of interacting elements to be processed simultaneously is reduced, imposing a lower load on working memory.

- *Extraneous cognitive load* is working memory processing that is ineffective or detrimental for learning. It can be associated, for example, with poor instructional presentation or design, such as redundancy of text, graphics, and spoken words. In this case, having the same information presented in three different formats may demand an extra processing that does not produce better cognitive schemata, and may take cognitive capacity away from the processing of relevant content.

- *Germane cognitive load* is cognitive load imposed by processes effective for learning. As stated by Wang and Antonenko, "germane load occurs when assimilation or accommodation of presented information is encouraged during learning challenging learners but not overwhelming them." ([**wang2017instructor**], page 81).

Accordingly, designers of multimedia instruction aim at creating materials that reduce extraneous cognitive load, leaving space for germane processing to happen.

The face of the instructor is a complex visual stimulus to be processed while learning. In the context of e-learning with videos that include instructor presence, it is necessary to acknowledge the importance this element has on learner's attention distribution, as instructor's face attracts significant attention (for revision of face role in human-computer interaction, see [**yee2007meta**]). When an instructor is present, the learner is presented with two sets of complex visual stimuli: the instructor image, and the rest of the video frame, that contains other instructional components, such as text, diagrams, images, *etc.*, including information relevant for learning. As both these complex sets of visual stimuli are processed by the visuospatial sketchpad, they compete for the same cognitive resources, potentially resulting in split attention [**van2017effects**, **kalyuga1999managing**, **antonenko2010influence**]. As the distribution of visual attention is impacted by the presence of the instructor face, attention may be given to the image of the instructor in detriment of the relevant content present in the rest of the video frame. Consequently, from a cognitive load perspective, instructor's presence may be extraneous information that would hinder learning.

Empirical evidence confirms that, when present, the instructor attracts significant amount of attention from learners [**kizilcec2014showing**, **wang2017instructor**]. Nevertheless, the results about the trade-off between costs and benefits of instructor presence are contradictory. Homer et al. [**homer2008effects**] confirmed the hypothesis that when the instructor is added to a video lecture, subjects self-report higher levels of cognitive load. To the contrary, Chen and Wu [**chen2015effects**] results show significantly lower self-reported cognitive load in the picture-in-picture video format compared to the voice-over type.

Learner control in multimedia instruction

Intuitively, allowing learners to have control over learning environment appeals as a solution to the limited capacity of working memory, because it gives learners the opportunity to adapt the learning material to their own cognitive needs [**tabbers2010learner**]. In the perspective of Cognitive Load Theory, such intuition is corroborated, that is, learner-paced systems are beneficial for learning compared to system-paced environments. Giving the control to learners may help them understand the content, as it supposedly decreases extraneous cognitive load and allows learners to take the time they need to engage in deeper processing [**mayer2001learning**]. In a similar vein, the *interactivity principle* states that people learn better from a multimedia instruction when they are able to control pace and order of the presentation [**mayer2003nine**]. Empirical evidence has verified the positive effect of learner control over multimedia instruction (e.g., [**tabbers2010learner**, **kuhl2014impact**]). Some studies have also identified that learner-paced condition increases significantly the instruction viewing time (by 60% in [**tabbers2010learner**], and by 100% in [**kuhl2014impact**]).

The present study involves both novice and expert subjects. In view of the possible cognitive overload and lack of motivation for novice subjects, we decided to include learner controls (pause, play, advance/next chapter, go back/previous chapter) across all experimental conditions. We expect that the availability of controls will induce learners to try to understand better the lecture, leveling their motivation towards the experiment. Such assumption is necessary for using self-reported mental effort as an index of cognitive load [**paas1992training**], which is applied in this study.

Attention dynamics: eye-tracking studies

Traditional methods to assess multimedia learning, such as questionnaires and tests, measure the results of subjects' learning process, from which researchers try to infer the mechanics of how people learn with multimedia instruction. On the other hand, the use of eye-tracking gives researchers more insights about the learning process itself, and not only its results. Learners' eyes remain fixated on what is being processed in their working memory (eye-mind hypothesis [**just1980theory**]), therefore eye-tracking data about fixations may shed light on learners' attentional dynamics, and consequently, on their learning process. Being a powerful tool, eye-tracking has been used in recent years to empirically study multimedia learning. According to review made by Wang and Antonenko [**wang2017instructor**], eye-tracking research has helped define some general findings: (a) there is a strong link between eye fixations and learning outcomes [**boucheix2010eye**], (b) visual cues guide learners' attention [**boucheix2010eye**, **de2010attention**], (c) visual attention is guided by prior knowledge [**canham2010effects**, **jarodzka2010eyes**], and (d) when learning from a multimedia animation, subjects spend more time reading printed text than looking at the graphics [**schmidt2010closer**]. Additionally, the use of eye-tracking could elucidate the split attention mechanism, as it provides information on how learners' gaze is distributed across pre-defined areas of interest (AOIs), such as instructor image and relevant content.

Despite these general findings, little eye-tracking research has been made in the direction of understanding how different elements of video lecture design affect visual attention. A few studies have found profound changes in watching behavior when the instructor is present in the video. Kizilcec et al. [**kizilcec2014showing**] identified that 41% of viewing time was spent looking at the face of the instructor. Louwerse et al. [**louwerse2009embodied**] detected that 56% of the time learners looked at the pedagogical agent in spite of its size of only one-fourth of the display.

This study is important because, despite research made on the topic, empirical evidence about multimedia instruction format effectiveness on learning is still inconclusive, and two conflicting theories (i.e., social-cue hypothesis and interference hypothesis) are accepted by the research community. In addition, to the best of our knowledge, there are no results available in the current literature discussing how different types of knowledge are affected by video instruction styles. Therefore, empirical research is still necessary to better understand the effectiveness of video lectures designs on learning, and to uncover how effects of video lecture design on learning may differ for different types of knowledge.

2.3 Effects of video lecture design on learning

The central question of the present study is:

Which is the most effective video lecture design for different types of knowledge?

Wherein effectiveness is twofold defined :

- *learner satisfaction* with his learning experience, which includes his perceived learning, and all his affective response to the video, and
- *learning outcome*, that is, how much the learner effectively learns from watching the lecture

The approach taken is looking at this question through the perspective of the learner, that is, any person who watches a video lecture. The Internet offers all types of knowledge in all styles and layouts. Hence, the continuity of learners in longer courses, such as MOOCs, depends strongly in how they feel about their experience. As detected by some studies (e.g., [**kizilcec2014showing**, **wang2017instructor**]), this is not necessarily related to how much they have actually learnt. Therefore we found essential to take learner satisfaction into account when investigating the effectiveness of video lecture designs.

Nevertheless we are still interested in the content really absorbed by the learner. An objective measure is still necessary to a complete assessment of the design's effectiveness. A learning outcome score can provide such an assessment. In order to have a better understanding of the mental processes of learners, the learning outcome will be evaluated in three different levels: recognition, retention, and transfer (detailed in section 3.4). This will provide a comparison with the findings of non-equivalency between perceived and actual learning, and even to go beyond most studies that have only considered one or two forms of learning (i.e. only recall or only recall and transfer).

We focused our research in two video lecture designs: a voice-over slides, and the same video lecture with the video of a professor overlaid (for more details, see section 3.3). Thus, this study has three main inquiries about the effects of the instructor presence on learning. Firstly, we seek to understand how the learning outcomes are affected by the two video lecture designs. In addition, we are interested in the impact of the presence of the instructor on learners' engagement, satisfaction, and perceptions about their learning experience. Finally, we also searched to understand the effects of instructor presence on learner's attention distribution and viewing behavior.

Engagement, satisfaction, and subjects' perceptions of learning experience

As a first scope of research, we focus on the affective dimensions of learning. We are interested in the influence of prior knowledge and instructor presence on learners' engagement, perceptions, and satisfaction. Under the social cue hypothesis, we hypothesize that:

H_1: Seeing the instructor has positive effects on learners' engagement.

Research on flow of students while learning articulate that learners with more *a priori* abilities in the content taught reach higher levels of flow, and, therefore, of engagement. Hence, we pose the following hypothesis:

H_2: Prior knowledge and engagement level are in a direct relationship.

Regarding the satisfaction facet of learning, we seek to understand which video lecture design makes learners more satisfied with learning, and also which one makes learners more willing to watch the following classes of the MOOC. Our hypothesis is:

H_3: Instructor presence will have a general positive effect on learners' satisfaction and willingness to continue following the MOOC courses, but these effects will be stronger on subjects from non-Informatics profiles.

Hypothesis four is based on two aspects. First, the positive socio-emotional effects of instructor presence stated by social agency theory. In addition, we believe learners from Informatics are used to, hence prefer not having the instructor visible, evidenced by the preference for board-centric course styles exhibited in technology courses [**santos2016speakers**], and on the writers personal experience in the Informatics field.

The affective response learners have to the video lecture design is an important aspect multimedia instructor designers must take into consideration when choosing which design to apply in a course. Thus, we are interested in verifying if our experiment will validate the hypothesis:

$\mathbf{H}_4$: Nonverbal cues provided by the video of the instructor (social cues) increase learners' affective response.

Still in the perceptual scope, empirical evidence shows that learners perception of their learning is influenced by instructor presence, but not necessarily corresponds to how much the have effectively learnt [**kizilcec2014showing**, **kizilcec2015instructor**]. Wherefore, this study is also applied to verifying the hypothesis:

$\mathbf{H}_5$: Perceived learning is higher when instructor video is visible *picture-in-picture*, and perceived learning does not reflect the real learning performance.

Learning outcomes

Secondly, we pose the research question (RQ):

$\mathbf{RQ}_6$: Does there exist a video lecture design that is more effective for learning?

This means we seek to understand in which condition learning outcomes are higher. As discussed in this section, there are two conflicting theories defended in the literature. On one hand, the social-cue hypothesis states that the presence of the instructor would benefit learning, due to the non-verbal cues provided by the image of the professor. On the other hand, the interference hypothesis states that the instructor presence adds extraneous cognitive load and, therefore, hinders learning. Due to the mixed evidence from other studies in this field, our goal is to determine if one of these two theories can be experimentally verified.

In addition, we are also interested in exploring this question for conceptual, factual, and procedural knowledge, that is, if *does there exist a video lecture design that is more effective for learning each type of knowledge*. Each type of knowledge has its own specificities related to the learning process, imposing probable different effects of video lecture design on learning. Choosing this approach, our results will be attached to a type of knowledge present in all fields instead of being attached to the field of the chosen video lecture. The distinction of effects of video lecture design on learning for different types of knowledge, therefore, enables us to provide better guidance to multimedia instruction designers.

Attention distribution and viewing behavior

Eye-tracking data allow us to have a wider view of learners experience while watching video lectures. Besides learners' perceptions, this study also takes advantage of the eye-tracking technology to observe learners attention distribution and viewing behavior. This approach offers the opportunity to cross self-reported and observational data, allowing the exploration of interacting phenomena between learners' perceptions and their behavior. In this sense, we want to verify the findings that instructor presence profoundly impacts learners' attention distribution. In addition, we seek to understand if such impact on attention distribution differs for different educational profiles (i.e., Informatics and non-Informatics), based on the above-mentioned preference of technology MOOCs for board-centric designs. Hence, we study the following hypothesis:

$\mathbf{H}_{7a}$: Instructor presence attracts a fair amount of learners' attention.

$\mathbf{H}_{7b}$: Instructor presence impact on learners' attention distribution is more relevant for subjects of Informatics profile compared to non-Informatics subjects.

In regards to viewing behavior, two hypothesis are posed, supported by results on the use of learners' controls. The first hypothesis concerns the impact of prior knowledge on viewing times. Empirical evidence found that learners with more prior knowledge have longer viewing times, because they revised the content more times [**tabbers2010learner**]. The prior knowledge on Informatics declared by people on the field, and people from other fields may differ in profoundness. Hence, we hypothesize that:

H$_8$: For each educational profile, learners with more prior knowledge spend more time viewing the video lecture.

The second hypothesis is:

H$_9$: The longer learners watch the video lecture, the higher are their learning performance.

We base hypothesis H$_9$ in the fact that the use of learner controls, and consequent longer viewing time, indicates higher motivation to understand the lecture content [**tabbers2010learner**]. In addition, the use of controls gives learners more time for deeper processing [**kuhl2014impact**], allowing a better quality of acquired knowledge structures, and, consequently probable higher learning performances.

The method adopted to answer this research question is to set up an experiment where people would watch a lecture in video and, then, express their satisfaction about the experience and solve questions about its content. To fully profit from this experiment, observational data will be collected, such as eye gaze, facial expressions, and body movements. A detailed exposition of this approach is found in the next chapter.

— 3 —

Method: user experiment

The impact of layout on the effectiveness of video lectures was evaluated through a user experiment using material provided by the Inria Learning lab. In this experiment, eye-gaze, face expression and posture were recorded of subjects watching tutorial material in the area of semantic web programming presented with different layout format. Each subject was then tested for recognition, retention, and transfer of the lecture material. This section presents details of this experimental protocol, including the experimental hypothesis and goals, measurements, tutorial material, and the experimental parameters. Results from the experiments are presented in chapter 4.

The goal of the experiment is to understand better subjects behaviors while watching video tutorials and the impact of video design on learning different types of knowledge. Our goal is to describe the learning experience and determine learning outcome as a function of video layout and course content. Our intention is to correlate subjective observations with objective measurements in order to complete the interpretation of results with causal explanations.

3.1 Participants

A total of 24 volunteers participated as subjects in the experiment. Subjects did not receive any gratification for their participation. All subjects were residents of the Grenoble area. Subjects were contacted through the mailing lists of the laboratories involved in this study (Inria, LIG, LaRAC) as well as by direct personal contact of the members of the team. All subjects had normal or corrected-to-normal vision, and the majority have French as their native language (22 out of 24). The level of education for subjects varied from University level (Baccalauréat+2 in France) to University professor with a doctorate.

Among the subjects, half (12) had some background in Informatics, that is, they had passed through formal studies in the domain or had a deep knowledge of its theoretical basis. The other half (12) came from other backgrounds. Regarding prior knowledge about Web of linked data and semantic Web, 6 participants declared they had a good level of knowledge about topics treated during the selected video lecture.

There were two eligibility criteria. The first criteria was mastery of the French language, imposed because the experiment was completely carried out in French. Preference was given to French native speakers to avoid adding the language or cultural background as other parameters to the experiment. Only 2 participants were non-native, however both mastered French due to living and working or studying in France for more than one year. The second eligibility criteria was to have completed at least one University level educational program. This criteria was chosen to reduce variability in the learning outcome results, as it was considered that the education level would influence people's capacity to learn the subjects considered for the experiment. The threshold for this criteria was set at the Bachelor's degree or a French equivalent level (e.g. BTS, Licence, *etc.*).

3.2 Setup

Figure 3.1 presents the experimental setup. The same setup, with a minor modification, was already validated in a study of chess players mental models developed in our group by Guntz et al. [**guntz2017multimodal**]. The hardware components are: a 23.8" Touch-Screen computer with Windows 10 64-bit as operational system, a Kinect 2.0 mounted 35 cm above the screen focusing on the learner, a 1080p Webcam for a frontal view, a Fovio Eye-Tracking bar, a JBL on-ear headphone for listening to the lecture, and two adjustable USB-LED for lighting condition control. The Touch-Screen was used in an attempt to capture more body movements from learners for future analysis. In order to guarantee the same sensor placement and orientation for all participants, a laser-cut wooden super-structure rigidly mounts the measuring equipment over the computer screen.

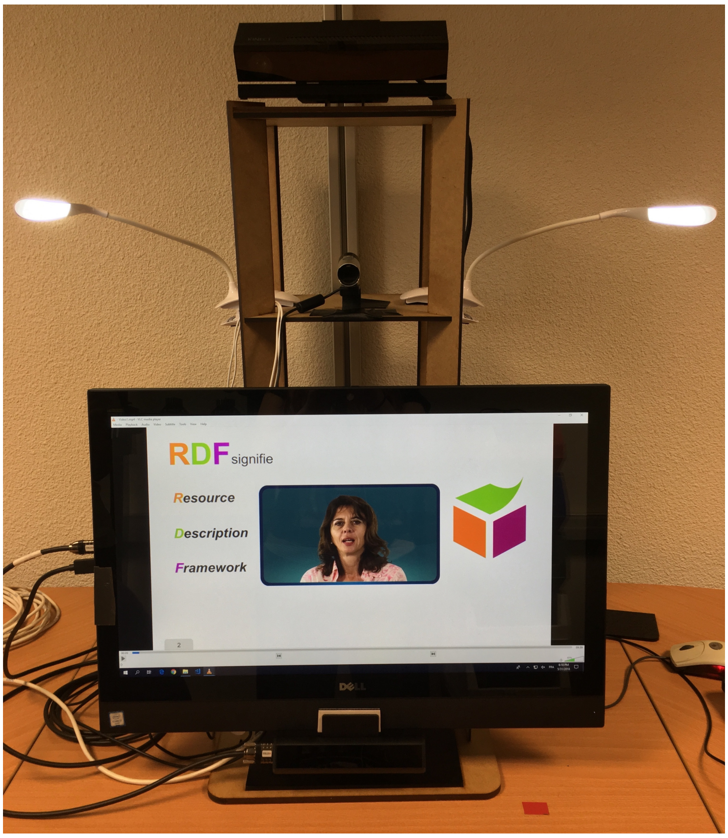


Figure 3.1 – Experimental setup

During the experiment, the subjects watched the video using the free and open source VLC multimedia player. Its interface configuration was personalized to have only the play/pause button, the previous and next chapters buttons, the volume button, and the time bar that allows subjects to move through the video lecture. Subjects were provided with headphones to help create a more absorbing learning experience, by isolating subjects from the environment and help them focus attention on the video lecture. The **EyeWorks™**[1] platform was used both during the experiment (***Collect*** solution) and during data analysis (***Analysis*** solution) together with the programming language R and **R Studio**.

[1] *EyeWorks*^{TM} platform: `http://www.eyetracking.com/Software/EyeWorks`

The data recording from all sensors (Screen capture, user clicks, Kinect 2, Fovio) was done using the RGBD Sync SDK2 from the MobilRGBD project [**vaufreydaz2014mobilergbd**]. The data from all sensors are synchronous thanks to the millisecond precision timestamp associated to each recorded frame. This framework also allows the displaying of all the gathered and computed data, as depicted in Figure 3.2.

Figure 3.2 – Multimodal recordings made by experimental setup. Left to right: RGB video of body posture with body joints, depth view from Kinect 2 sensors, screen recording (green point is current position of gaze), plot of current emotional valence (level of positive emotion), and frontal video with facial keypoints (white dots).

3.3 Video lecture

The choice of the video lecture has a key role in defining which parameters and behaviors will be possible to be observed and analyzed. This section aims at elucidating the options made and the criteria employed regarding the choice of this central element of the experiment.

Choice of subject

A MOOC about the **Web of linked data and semantic Web**[2] was selected as the resource for the video lectures used in this study. This MOOC is proposed by Inria's Learning Lab[3], group that offered us access to its repository for purposes of this study. We searched for a topic where we could easily compare the impact of layout on learning for factual, conceptual, and procedural knowledge. We considered the domains of Mathematics, Statistics, and Teaching a Programming Language. An additional criteria was the availability of MOOC materials in French for each kind of knowledge and each presentation design. In view of these criteria, the most appropriate subject was found to be programming the semantic Web.

Video selection

The video selection considered that all three types of knowledge should be addressed during class and should be clearly distinguishable, that both people with and with no background in Informatics should be able to learn something, and, at last, that the final duration would be 10 minutes maximum. Hence, two classes were chosen. The first, from the second week of

[2] Course available in the Fun MOOC platform: `https://www.fun-mooc.fr/courses/course-v1:inria+41002+session03/about`

[3] Inria Learning Lab: `https://learninglab.inria.fr/en/`

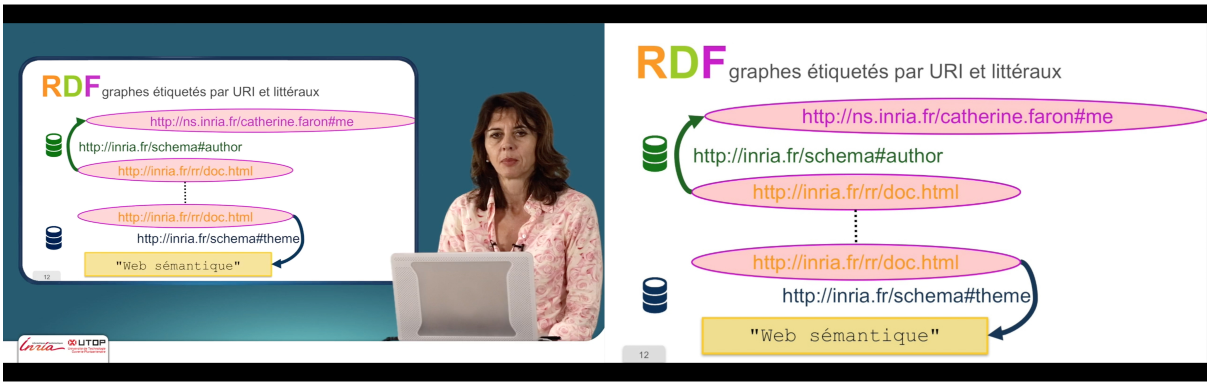


Figure 3.3 – Two experimental conditions. Left: video lecture design *with instructor*, that is, slides overlaid by picture-in-picture video of the instructor narrator. Right: video lecture design *without instructor*, that is, voice-over and slides.

the course, addressed some fundamentals of the Web of linked data, such as RDF (Resource Description Framework). The second, from the week 6 out of 7, presented one application of the Web of linked data: the vocabulary Friend of a Friend (FoaF). The two videos of the first class and the video of the second one were combined and edited with iMovie and Final Cut Pro. The head of the Inria Learning Lab, Dr. Jean-Marc Hasenfratz assisted us in putting this material together.

This MOOC employed the video design were the instructor is present in overlapping the learning material, and in a non-fixed position. From the slides of the course, we were able to generate a second version were the video design was voice over slides. Therefore, there were *two final versions* of the video lecture, both lasting, approximately, 10 minutes: *one with the presence of the instructor and one without it*. A sample of each version is show in Figure 3.3.

Chapters and classification in types of knowledge

In order to conduct an analysis based on types of knowledge, the video was sliced in 16 chapters and the type of knowledge addressed in each one was defined. Two people watched the video and classified the chapters according to the definitions of types of knowledge provided in Section 2.1. Then, for all the different labelling, a consensus was reached between these two people. The results are shown in Table 3.1.

The chapters were defined based on a tree of the lecture's content (see Appendix A) we built from our understanding of such lecture. This tree also helped in the type of knowledge classification, and, later, in the choice of which elements of the lecture should be measured to calculate the learning outcome score.

3.4 Measures

In this study both objective and subjective measures were gathered. Both are complementary and will allow a richer understanding of the learning process when watching a video lecture.

Chapter	Chapter name	Type of knowledge	Start [sec]	End [sec]	Duration [sec]
1	RDF: idea	conceptual	0	9	9
2	RDF: definition	factual	9	44	35
3	RDF: triplets	conceptual	44	53	9
4	RDF: example	procedural	53	102	49
5	RDF: atom of knowledge	factual	102	113	11
6	RDF: from triplet to graph	procedural	127	163	36
7	RDF: graphs	conceptual	163	210	47
8	URI and literals	factual	210	242	32
9	Linked data	factual	242	272	30
10	Connected graphs	conceptual	272	315	43
11	Views on graph	procedural	315	355	40
12	FoaF	factual	355	467	112
13	FoaF example description)	procedural	467	485	18
14	Related vocabularies	factual	485	545	60
15	Relationship	factual	545	569	24
16	Sub-properties of *foaf: knows*	procedural	569	592	23

Table 3.1 – Chapters of the chosen video lecture classified by type of knowledge presented (cf [**anderson2001taxonomy**]).

Multimodal observation

The sensors described in Section 3.2 provide objective multimodal observational data. Our goal is to extract from them a diverse range of measures that allow us to characterize the viewing behavior of learners, as done for Chess players in Guntz et al.'s study [**guntz2017multimodal**].

Eye gaze The eye-tracking device collects the subjects' eye gaze, and pupil dilation. From this data we can extract, for example, the attention distribution, that is, how long did participants spend looking at the instructor. This can be further compared with previous results from the literature (van Gog [**van2017effects**]). The pupil dilation can be used to estimate the cognitive load imposed by the video lecture to the learner. Combining both and the learning outcome scores helps us determine if the instructors' presence served as germane or extraneous cognitive load, that is, helped or hindered the learning process.

Body posture The body posture of each subject is computed from the video recorded by the Kinect. From this data, two measures are computed: the self-touching, collisions between wrist-elbow segments and the head (see [**aigrain2016multimodal**]), and body volume, space occupied by 3D box built around joints (see [**johal2015cognitive**]). They can be used as an indication of negative affect and frustration, that could come from a too high cognitive load and be a sign of the presence of extraneous load.

Affect The Webcam video provides us with the facial expressions of learners during the video lecture. The analysis of their micro-expressions (Ekman and Friesen [**ekman1969nonverbal**]) can be crossed with the self-reported affective measures in order to understand the learners satisfaction with their learning experience.

Self-reported measures

Are the videos interesting and easy to follow? Do I feel I have learnt a lot? Does the instructor look nice? These are all questions we have asked ourselves in order to decide whether to continue or not following an online course. This leads us to believe that the engagement and retention of learners in a MOOC are strongly guided by their feelings about and their satisfaction with the course. Therefore, tracking these variables is really important to define the effectiveness of a video lecture. To do that, a post-questionnaire was applied just after participants finished watching the video (see Appendix D for the full questionnaire).

To assess learners' engagement, the Flow in Education Scale (EduFlow Scale) [**heutte2016eduflow**] was applied. This scale is an adaptation made by Heutte et al. developed to study how learning environments support learners' flow. Flow, or autotelic experience, is defined as the "the experience of complete absorption in the present moment" (Nakamura and Csikszentmihalyi 2009, p. 195). From the four dimensions of EduFlow, only two were considered coherent with the context of this experiment:

- Time transformation (Flow D2) - alteration in the perception of time, sometimes leading to a lengthened duration of immersion in the task, and
- Autotelic experience (Flow D4) - well-being during task performance resulting from purpose in the task itself that enhances persistence and desire to engage in the activity again.

The loss of self-consciousness (Flow D3) and the cognitive absorption (Flow D1) were not included because we judged it would be more appropriate for more active tasks during learning. The proposed experience of watching passively the video lecture was thought not to provide enough action, and, therefore, participants would have difficulties measuring them.

The post-questionnaire also contained questions regarding the perceived learning, the learners' self-reported satisfaction, their self-reported affective response, and the self-reported mental effort it took them to watch the video lecture. These measures were inspired by similar studies present in this field (Wang and Antonenko 2017 [**wang2017instructor**], Kizilcec 2014 [**kizilcec2014showing**]). The statement for the question about the affective response to the instructor presence made explicit, for each experimental condition, the nature of such presence: either by its discourse or by its discourse and its image. Thus, there were two versions of the post-questionnaire, one adapted to each experimental condition.

After solving the learning outcome test, participants were asked once again to estimate their mental effort. But this time it was the mental effort put into solving the questions in the learning outcome test. If subjects are considered to have the same level of motivation and engagement, the mental effort can be considered an index of cognitive load [**paas1992training**]. Paas and van Merriënboer developed a measure of instructional efficiency that takes into account the mental effort put into solving the test and the performance in that test ([**paas1993efficiency**], revised in [**van2008instructional**]). An adapted version of the instructional efficiency measure was widely applied in others studies, where subjects were asked to evaluate the mental effort put into watching the lecture, and this was compared to the learning outcomes. The original measure evaluates the learning *outcomes*, giving an indication of the quality of cognitive schemata acquired by the learner. On the other hand, the adapted measure evaluates the learning *process*. We chose to ask subjects to measure both mental effort while watching and while solving the questions, allowing a richer analysis through the computation of both original and adapted instructional efficiency measures.

After the participants finished solving the learning test, we asked them about their sensations and impressions about the experiment. The post-questionnaire gave an overview of learners' affect, effort, and perceived learning for the lecture as a whole. So, during this conversation, in order to qualitatively enrich the questionnaire results, we went back to its questions and asked the participants if they had anything to add. In an attempt to differentiate these measurement per type of knowledge, we also asked them if the experience they described was the same during the entire experiment or if it was different for some parts.

Learning outcome

The test applied to the participants provides objective measures of their learning (*cf.* Appendix E). This learning was measured in three different levels: recognition, retention, and transfer. Having all these levels will provide a broader understanding of the effects of video design in the subjects' learning process.

One challenge while building the learning test was to measure all types of knowledge in all three levels in a balanced way. Another one was to build statements and put them in an order that did not guide subjects' memories into artificially remembering content that they would not have remembered by themselves. A solution found was to never address the exact same topic twice. A pool of topics was selected for each type of knowledge, and then distributed evenly throughout the different sections of the test. The transfer part, however, by its nature, assesses various topics and the video content as a whole.

Recognition When presented with a list of topic, recognition is the ability to determine which topics that were and were not addressed during the lecture. Conceptual knowledge was tested with recognition of 6 items, and both factual and procedural knowledge were tested using 4 items each. These items were organized in two questions (*cf.* questions 1 and 2 in Appendix E): one testing recognition of items that were present (conceptual), and the other testing recognition of items that were not present (factual and procedural).

Retention Retention is related to the process of retrieval of information from the past. However, retention goes one step further than recognition. It includes remembering finer details about what was learnt, with an ability to describe what was presented. It is closely related to reproducing what was learnt but it does not comprehend the ability to articulate it into new knowledge. Seven questions were applied to test subjects' retention (*cf.* questions from 2 to 9 in Appendix E): 4 multiple choice (MC) and 3 open questions. The option IDK - "I don't know" (in French JNSP - "Je ne sais pas") was added to the MC questions in order to reduce variance from guessing. Participants were also instructed to leave in blank any of the open questions they did not know how to answer. These cases were also marked as IDK in the dataset.

Transfer Transfer refers to the ability of applying what was learnt to a similar but not identical situation or problem. Here we applied a near transfer test (*cf.* question 10 in Appendix E), where the subject was asked to solve a problem similar to what was viewed in the course, but that demanded a more complete understanding of the content. For that the participant would need to have built a mental model of what was taught during the lecture, for example, being able to make significant connections between aspects treated in different parts of the video. The quality of this mental model is also influenced by the prior knowledge in the subject. This is one of the reasons why a measure of prior knowledge was added to the pre-questionnaire. For the transfer question participants were also instructed not to guess and to leave it blank in case they did not know really how to solve it.

Due to the time constraint and the richness of the collected data, not all these measures will be explored. However, the data will be made available as open data. So giving the overview of the possible measures to be extracted from this data contributes to other studies that may want to explore our dataset.

3.5 Experimental protocol

Experimental protocol stands for how all the measures and aspects of the experiment explained in the previous sections were organized when the experiment was carried out. A preliminary experiment was conducted with 4 participants (2 informatics, 2 non-Informatics) in order to validate the measures and the experimental protocol. This test phase allowed us to correct some phrasing errors found in the questionnaires and, more importantly, showed us that the learning assessment test took to long to be solved, mostly by non-Informatics subjects. Therefore, from the 2 questions that had been formulated for the transfer test, only one was maintained, and, in addition, it was simplified by adding an example of what was expected as an answer. Finally, after these modifications, the experiment consisted of **six steps**:

1. **Greetings and experiment overview**

   The participant was thanked by his or hers voluntary participation and was asked to sign a consent form (cf. Appendix B), imposed by the collection of personal data. Then, the experiment steps where laid out, and time for questions was given. After all doubts had been solved, the participant was given the pre-questionnaire.

2. **Pre-questionnaire**

   The pre-questionnaire (full version in Appendix C demanded the participant's education background, education level, and prior knowledge, in this order.

3. **Watch the video lecture**

   Firstly, the user controls available were explained (pause, previous chapter, next chapter). A general explanation about Web of linked data and semantic Web was, then, given to the participant, who solved occasional doubts. Next, the calibration of the eye-tracking device was made. Each participant was given a maximum of 15 minutes to watch the lecture, that is, 50% extra its duration, in order to be able to explore the user controls given. A 9-point calibration was done for adapting the eye-tracking to each participant.

Right after, a timer of 15 minutes was set, the participant put the headphones on, and started watching the video. Participants did not take notes while watching the video. We went, then, to a different table in the same office, far from the experimental table (to leave the participant at ease), and the experimenter waited until the participant had finished watching the video.

4. **Post-questionnaire**

   The participant was asked to first focus on his feelings about the video lecture experience. The post-questionnaire demanded the participants to fill out the Edu-Flow scale, and to report their perceived learning, satisfaction, mental load, and affective response, in this order.

5. **Learning assessment**

   The participant solved the questions of recognition, retention, and transfer, in this order.

6. **Conversation about general remarks** In the end, as explained in Subsection 3.4, the participant had space to express any general remarks about the experiment, or opinions about improvements to be made in the protocol. Then, the participant was asked about his or hers education background, to refine the information given in the pre-questionnaire. At last, focus was put back to the post-questionnaire and the learning assessment. The participant was asked if he or she had something to add about any of the questions. Participants were specifically asked to comment more about the questions of affective response to the instructors' presence.

Before leaving, participants were thanked again and offered cookies and juice. In addition, they were offered to receive the results of the study with some global explanations.

— 4 —
Results

In seeking to find which lecture design is more effective for learning, we performed a pilot experiment, detailed in chapter 3. This section presents a first analysis on the data collected from this pilot experiment. Firstly, a summary of the data collected is provided, followed by a description of how the data was prepared to be analyzed (e.g., grading criteria for learning assessment, eye-tracking data selection). This chapter also presents the statistical and qualitative results for the research question about learning outcomes (RQ_6), and for the hypothesis about engagement, satisfaction and perceptions (H_1, H_2, H_3, H_4, H_5), and attention distribution and viewing behavior (H_7, H_8, and H_9; See section 2.3).

4.1 Collected data

Each of the 24 participants answered a total of 21 questions that provided 38 data points per participant, i.e., a total of 912 data points collected from subjects' questionnaires (pre-questionnaire, post-questionnaire, learning assessment). These points are divided into 3 profiling questions, 8 self-reported measures, and 10 learning assessment tests. The learning assessments were divided into 3 types of learning outcomes: recognition (2 questions), retention (7 questions), and transfer (1 question). Regarding observational data, approximately 35 Gb of data was recorded for each participant. Observational data included environment audio, facial video (video camera), body video (Kinect), and eye-tracking data. No subject was excluded from analysis, and all participants have valid data for every measure taken.

Subjects are divided into 4 groups according to 2 control variables: experimental condition, and profile. Experimental condition is the video lecture design of the video watched by the subject, that is, either *with instructor* or *without instructor* (see Figure 3.3). Subject's profile, i.e., informatics or non-informatics, is based on the profiling questions, and was confirmed during the final conversation with the participant. Subjects were randomly assigned to one of the two experimental conditions, creating 4 groups of 6 subjects each. Subjects' prior knowledge did not significantly differ across experimental conditions ($M_{with} = 1.60$, $M_{without} = 1.43$ out of 5 on the prior knowledge questionnaire; Wilcoxon test: $W = 81, p = 0.616$).

Viewing time of each subject was extracted from the eye-tracking recordings. In order to analyze subjects' attention distribution, the video lecture was divided into scenes, i.e., a video span showing the same slide and the instructor in the same position, if applicable. Each scene is part of a chapter, and, consequently, is related to a type of knowledge. The video lecture includes 36 slides. Hence, the *without instructor* condition has 37 scenes, as one of the slides is used in two chapters. In the *with instructor* condition, the instructor video is not always in the same positions, and, at times it is not present. Therefore, 52 scenes were identified, from which 41 included the video of the instructor. In each scene of the *with instructor* condition, regions delimitate content and the instructor area, allowing a deeper analysis of the effects of instructor presence on subjects' attention distribution.

4.2 Data analysis

The experiment performed provides a substantial amount of data, ranging from subjects' perceptions of their learning experience, to measures of their viewing behavior and effective learning. Before analysis, the paper questionnaires were digitized and organised in order to provide coherent data for analysis. This section presents the criteria used in the correction of the learning assessments, and the method applied to transform the raw data points into the measurements analyzed in the results.

Data preparation

Learning outcome scores were computed from the grading of subjects' learning assessment test. The test is divided in recognition, retention, and transfer, and covers a set of pre-defined key topics of the lecture content, balanced across the conceptual, factual and procedural knowledge. To calculate recognition scores, each of its 14 items was given a value of 1 in case it was correctly selected, and zero otherwise. Therefore, each assessed topic have the same weight. Retention questions assess only one topic each, however, their solution may include one or more elements. Each element was given one point if correct and zero if wrong or missing. In order to give the same weight to each evaluated topic, questions scores were normalized, and only then summed. The transfer question assessed subjects' ability of applying what they just had learnt. Its solution included 8 RDF graph elements and 3 elements adjacent to the RDF graph. A correctly drawn element was awarded one point, and a missing or wrong element was given zero point. Half a point was awarded in three situations: a graph element correctly placed but with no label on the arrow, a graph element correctly written but without its identifying character, or when the arrow was correctly labeled but in the wrong direction. Recognition, retention, and transfer scores are the normalization of the sum of points of the respective questions. Learning outcome scores are the sum of recognition, retention, and transfer scores, therefore, varying from 0 to 3 points.

Engagement and subjects' perceptions were evaluated through eight self-reported measures: EduFlow scale, perceived learning, satisfaction, continuity in MOOC (i.e., willingness to watch the next videos of the same MOOC), mental effort while watching video lecture, affective response to presence of the instructor, and mental effort while solving the test. The measures with more than one item were represented by the mean and standard deviation of its items values. In the case of EduFlow, two dimensions of engagement were measured, each one contributing separately to the final flow effect. Therefore, each dimension score is the mean of its three items, and the total score is the sum of the dimensions values.

Each eye-tracking recording provided a video of the display overlaid by the a green dot representing subject's gaze, plus the coordinates and durations of subjects' gaze. In order to run analysis over this data, each video needed to be divided into the pre-defined scenes. So, for each subject, each scene start and end time was determined with frame precision. Scene identifiers are formed by its number (1 to 37 in *without instructor* condition, and 1 to 52 in *with instructor* condition), the chapter number, the slide number, and, the number of the scene in that chapter, when applicable. Due to learner controls, one subject may have visited the same scene more than once. In these cases, a dot notation position identifier was given to the scene. For example, if the scene 3 was watched for the second time, the identifier would be *03.1_ch2_sl3_sc2*, that reads, scene 3, chapter 2, slide 3, scene 2 of chapter 2.

Data treatment

The data collected is organized into two main datasets: one for the data points collected on paper and transcribed into *csv* files by the research team, and one for subjects' viewing times based on the eye-tracking data. These datasets contain the raw data. After the preparation, both datasets were cleaned, and indexed by subject id (from 1 to 24) in order to have an anonymized dataset. Due to same index column, data from both datasets can be simultaneously analyzed. The observations of facial expressions and body posture were not treated in this first analysis, however the raw recordings will be made available together with all collected data as open data.

A more intensive treatment was given to the eye-tracking data. As described in the previous section, we had the gaze position and duration for each subject during each scene. Using the Eye Works^{TM} Analysis software, each scene was associated to an image depicting the video lecture at that point. The superposition of the gaze information over the scene image allowed a better understanding of subjects' viewing behavior. As the video lecture design was different in each experimental condition, the analysis of subjects' gaze was also made separately. The *with instructor* condition allowed us to explore the impact of instructor presence on subjects' attention. In order to collected numerical data about it, each scene image had content and instructor regions delimited. The number of eye fixations in each region was then extracted. The minimum length of fixation was set to 75 ms. A similar exploration of region fixations could be made for subjects' in the *without instructor* condition, in order to understand how long do learners look at the content.

4.3 Results

Four groups are considered in the analysis, i.e., combinations of the two experimental conditions (*with instructor*, and *without instructor*) and the two educational profiles considered (Informatics, and non-Informatics). All four groups have independent samples, as each subject belongs to only one of them. In addition, in this study, the significance level considered is $\alpha = 0.5$.

For comparison of means between two independent groups, unpaired two-samples t-tests are applied. In all cases, the normality assumption of the t-test is verified using a Shapiro-Wilk normality test, and F-tests were applied to test for homogeneity in variances. In cases of non compliance with the t-test assumptions, an unpaired two-samples Wilcoxon test is performed instead. For tests of one mean is significantly greater or less than the other, the one-tailed version of the statistical test is employed. In cases where the significance of the difference is in check, the two-tailed version of the statistical test is applied.

Engagement, satisfaction, and subjects' perceptions of learning experience

Subjects' engagement was measured using the EduFlow scale (cf section 3.4). EduFlow scores were compared between the two experimental conditions: $M_{with} = 7.25(1.76)$, $M_{without} = 6.75(2.67)$. Even though subjects in the *with instructor* condition presented higher engagement scores, these scores are not significantly higher than the engagement scores for *without instructor* subjects ($t = 0.5416, p = 0.297$). Therefore, the hypothesis H_1 is not validated by the results, i.e., seeing the instructor does not have positive effects on learners' engagement.

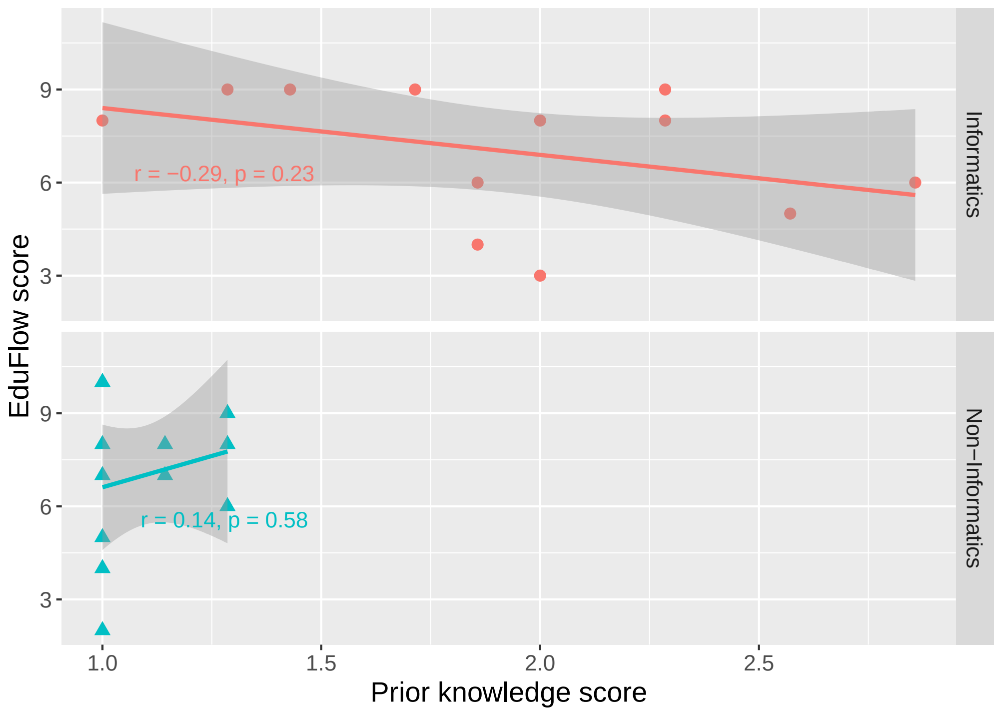


Figure 4.1 – Relationship between prior knowledge scores and engagement, separated by subjects' educational profile (top: Informatics profile; bottom: non-Informatics profile). Regression lines are shown with a 95% confidence region. The colors and shapes are used simply as a visual aid.

The relationship between prior knowledge and subjects' engagement was also inspected (Figure 4.1). The video lecture treats a topic in Informatics, imposing different interpretations of subjects' prior knowledge according to their educational profile. Therefore, this relationship was inspected separately per subjects' educational profile. For the non-Informatics profile, the results show a non significant slight direct relationship between prior knowledge and engagement scores (Kendall's tau coefficient: $\tau = 0.14, p > 0.1$). For the Informatics profile, subjects' prior knowledge and engagement are inversely related, what would refute the hypothesis H_2. However, in the Informatics case, the correlation is also not significant (Kendall's tau coefficient: $\tau = -0.29, p > 0.1$), therefore this analysis is neither able to confirm nor to reject the hypothesis H_2.

Attention was given to the effects of instructor presence on learners' satisfaction and willingness to continue following the MOOC classes (i.e. continuity score), as shown in Figure 4.2. In terms of satisfaction, instructor presence tend to have a negative effect ($M_{with} = 5.00(1.41)$, $M_{without} = 5.83(1.19)$; Wilcoxon test: $W = 45.5, p < 0.1$). Learners continuity in MOOC suffered no impact from experimental condition, showing the same average for the two experimental conditions ($M_{with} = 5.42(1.43)$, $M_{without} = 5.42(2.23)$; Wilcoxon test: $W = 64, p = 0.661$). Therefore, the presence of an instructor had no significant effects on learners' continuity in MOOC, and tended to have general negative effects on learners' satisfaction, refuting the stated on hypothesis H_3 for satisfaction.

The verification of a stronger positive effect of instructor presence for non-Informatics subjects was done in two steps. First, it was evaluated if, for the condition *with instructor*, subjects from non-Informatics profile are more satisfied and willing to continue the course

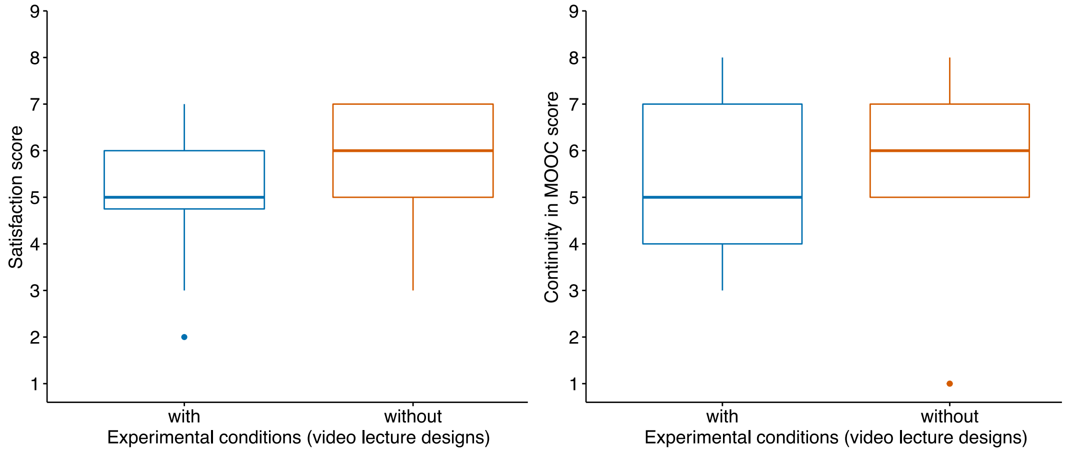


Figure 4.2 – General effects of instructor presence on subjects' satisfaction (left), and willingness to continue following the MOOC classes (right).

than Informatics subjects. This effect is supported by the results for continuity in MOOC, but not for satisfaction (Continuity score: $M_{with} = 4.00(0.63)$, $M_{without} = 6.83(1.17)$; t-test: $t = -5.22, p < 0.001$). The second step verified the interaction between the two controlled factor, i.e, educational profile and experimental condition, in regards to satisfaction and continuity scores. A factorial ANOVA was performed for each of the scores. Figure 4.3 represents the interaction of the control factors in both situations. Satisfaction scores were not affected by an interaction of both factors. On the other hand, the relationship between continuity scores and experimental condition tend to depend on educational profile (F$= 3.351$, $p = 0.082$).

The affective response of subjects to the presence of the instructor was analyzed firstly through a comparison of means (Figure 4.4). The ratings given to the five adjectives describing the instructor presence were averaged into one affect score, ranging from 1 to 9. Subjects in the *with instructor* condition reported a significantly lower affective response compared to subjects' watching the voice-over design ($M_{with} = 6.37(1.81)$, $M_{without} = 8.43(0.98)$; t-test: $t = -3.47, p = 0.001$). The results not only provide evidence to reject the proposed hypothesis

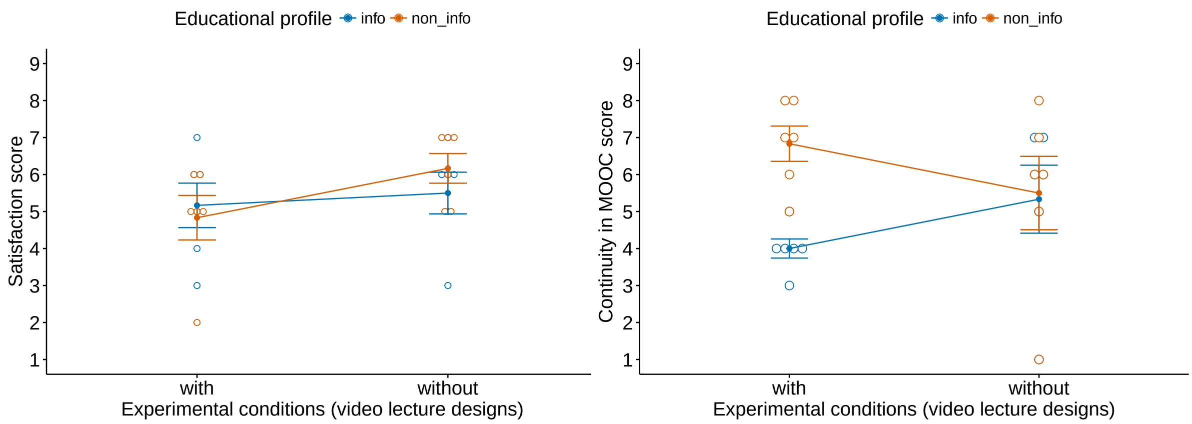


Figure 4.3 – Interaction between experimental condition and educational profiles over satisfaction (left) and continuity in MOOC (right) scores.

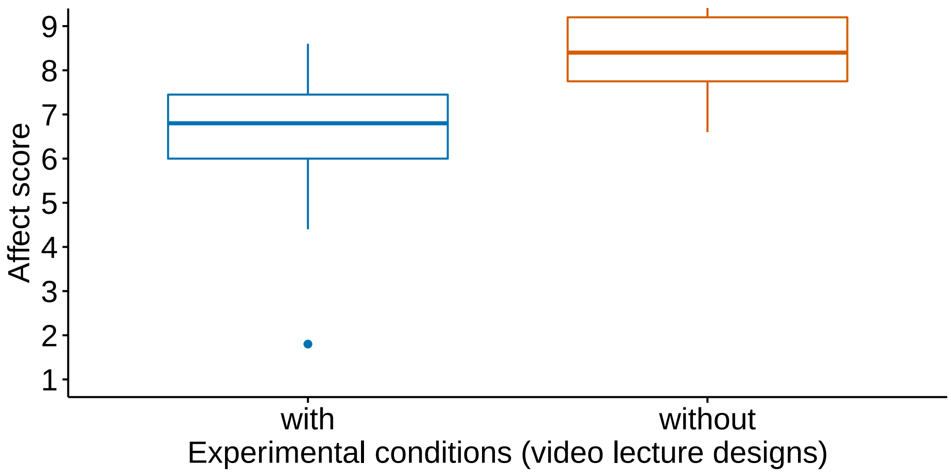


Figure 4.4 – Affective response of subjects to the two video lecture designs, i.e., with instructor visible, and just voice-over slides.

H_4, but also indicate the opposite phenomenon, i.e., instructor presence has a negative effect on subjects' affective response. Furthermore, a MANOVA analysis was performed in order to understand the influence of each of the five adjectives (i.e., "aidante", "utile", "frustrante", "gênante", "confuse") on affective score. Only "aidante" ($p < 0.001$) and "utile" ($p < 0.01$) significantly characterize subjects' affective response.

Effects of instructor presence on perceived learning were verified using a mean comparison. Subjects in the *with instructor* condition perceived having learnt significantly less than subjects in the *without instructor* condition ($M_{with} = 4.67(1.61)$, $M_{without} = 5.83(1.34)$; *t*-test: $t = -1.93, p = 0.033$). Therefore, the observed effect is the opposite of the one hypothesized in the first part of H_5. Perceived learning was found not to be significantly correlated to total learning outcomes (Kendall's tau coefficient: $\tau = 0.24, p = 0.126$), as illustrated in Figure 4.5. However, comments of non-Informatics subjects in the end conversation session of the exper-

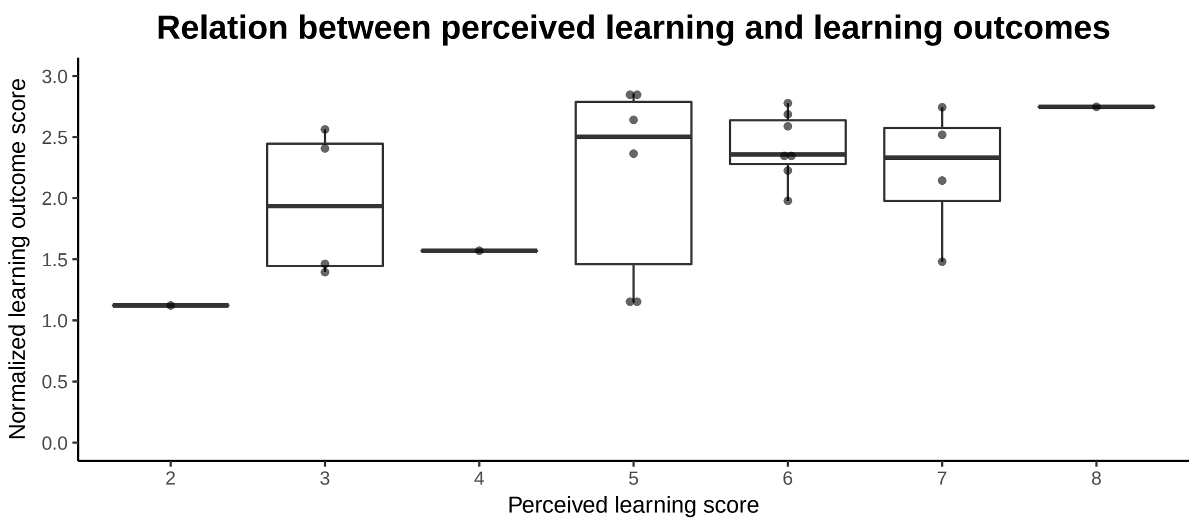


Figure 4.5 – Relationship between perceived learning and learning outcomes (n=24).

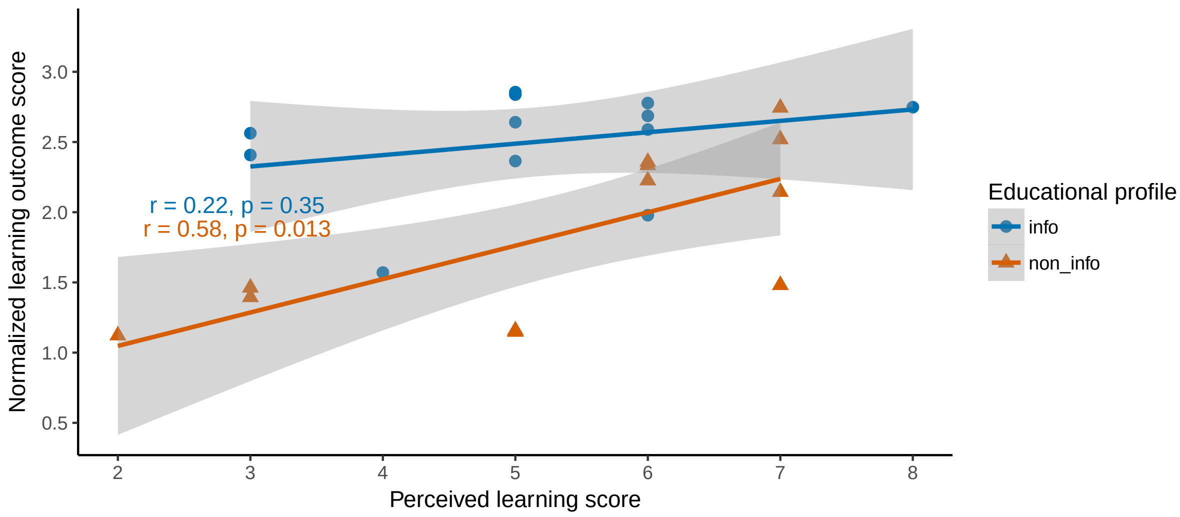


Figure 4.6 – Relationship between perceived learning and learning outcomes discriminated by educational profile (n=12). Regression lines are shown with a 95% confidence region.

iment informed us that many of them felt they had learnt a lot, because it was a new topic, even if they were not able to answer many questions. Therefore, it was decided to look at this relationship in taken into account subjects' educational profile (Figure 4.6). Results show that, while Informatics profile did not exhibit significant relationship between perceived learning and learning outcome, a significant positive correlation related perceived learning to learning outcomes for non-Informatics subjects (Kendall's tau coefficient: $\tau = 0.58, p = 0.013$). The final result is that the second part of hypothesis H_5 is supported when considering all participants together, but it is rejected for subjects of non-Informatics profile.

Learning outcomes

Effectiveness of learning takes into account only the learning assessment test scores. The recognition, retention, transfer, and total scores for each experimental condition can be found in Table 4.1. The *with instructor* condition presented higher recognition scores. For all the other learning outcome levels, i.e., total, retention, and transfer, subjects in the *without instructor* condition performed slightly better. However, as no significant difference between learning scores means was found across experimental conditions, the social-cue hypothesis and the interference hypothesis are neither confirmed nor denied by the collected data.

The effectiveness of video lecture design for learning is also explored per educational profile

Table 4.1 – Mean and standard deviation values of total outcome, recognition, retention, and transfer scores for the two experimental conditions (n=12). Total score is the sum of recognition, retention, and transfer scores.

Experimental condition	Total [0 - 3]	Recognition [0 - 1]	Retention [0 - 1]	Transfer [0 - 1]
With instructor	2.11 (0.65)	0.87 (0.12)	0.59 (0.24)	0.65 (0.39)
Without instructor	2.24 (0.56)	0.81 (0.11)	0.72 (0.15)	0.71 (0.36)

Table 4.2 – Mean and standard deviation values of total outcomes, recognition (recog), retention, and transfer scores for the four groups (n=6). Total score is the sum of recognition, retention, and transfer scores.

Educational Profile	Experimental condition	Total [0 - 3]	Recog [0 - 1]	Retention [0 - 1]	Transfer [0 - 1]
Informatics	With instructor	2.52 (0.49)	0.91 (0.13)	0.74 (0.20)	0.87 (0.19)
	Without instructor	2.48 (0.28)	0.86 (0.11)	0.78 (0.04)	0.84 (0.18)
Non-Informatics	With instructor	1.69 (0.52)	0.83 (0.10)	0.44 (0.19) *	0.42 (0.42)
	Without instructor	1.99 (0.68)	0.76 (0.10)	0.65 (0.20) *	0.58 (0.46)

Informatics profile: no significant differences between score means - total: $W = 25, p = 0.298$; recognition: $W = 23.5, p = 0.415$; retention: $W = 16, p = 0.810$; transfer: $W = 21.5, p = 0.628$. Non-Informatics profile: retention scores tend to differ ($t = -1.90, p = 0.087$); the other score means show no significant difference - total: $t = -0.85, p = 0.416$; recognition: $t = -1.23, p = 0.247$; transfer: $W = 15.5, p = 0.747$
* *tendency to differ, i.e.,* $0.05 < p < 0.1$

(see Table 4.2). Each educational profile (i.e., Informatics and non-Informatics) was analyzed separately. Learning outcomes of subjects from the non-Informatics profile show the same pattern as the general results, that is, apart from recognition, the group without the instructor presence presented slightly higher scores (differences not statistically significant). But, in this case, the differences are more accentuated than in the general case, with a 0.30 difference on total outcome score for the non-Informatics sample instead of a 0.13 difference for the all participants sample. Subjects from the Informatics profile showed a different pattern, with the *with instructor* group presenting slightly better global performance, even though the difference was not statistically significant.

All learning scores of the Informatics profile are not significantly different across experimental conditions (p < 0.05). For the non-Informatics subjects sample, the total outcome, recognition and transfer scores also did not differ significantly across experimental conditions. However, retention scores of subjects from non-Informatics educational profile tend to differ for different video lecture designs ($M_{with} = 0.44$, $M_{without} = 0.65$, $t = -1.8993, p = 0.08672$). Therefore, the results remain inconclusive in terms of verifying the social-cue and interference hypothesis.

Attention distribution and viewing behavior

The video of the professor attracted, in average, 11.59% of subjects' recorded eye fixations (i.e., eye gazes of minimum 75 ms with a threshold of 5 pixels). Subjects from Informatics looked more at the instructor ($M_{info} = 12.31(6.53)\%$, $M_{non\text{-}info} = 10.88(4.89)\%$; one-tailed t-test: $t = 0.43, p = 0.339$). Although no statistically different, these results indicate a possible opposite effect compared to the board-centric preference of Informatics students. However, this impact of instructor presence on Informatics' subjects distribution of attention is not significant, imposing a non-conclusive result about educational profile influence on attention distribution.

A qualitative analysis was carried out to study how instructor presence impacts attention distribution for each type of knowledge Figure 4.7 compares the heat maps of three video lecture scenes, one presenting factual (top), another conceptual (middle), and the third procedural knowledge (bottom). The scenes chosen presented the highest percentage of fixations on the professor per type of knowledge. The image of the instructor takes a fair portion of subjects' fixations for all three scenes (from 12.5% to 35.6% of all fixations over lecture content and professor). An interesting phenomenon is that for procedural knowledge, the attention given to the instructor image is less accentuated, maybe influenced by the greater amount of relevant content displayed. In summary, instructor presence has a fair impact on subjects' attention distribution, and this effect differs for conceptual, factual, and procedural knowledge.

Figure 4.7 – Heat maps of subjects fixations while watching the video lecture. For each type of knowledge, the scene with higher intensity of fixations on instructor was selected, allowing a fairer comparison. The scenes depicted presented samples of factual knowledge (top), conceptual knowledge (middle), and procedural knowledge (bottom). The percentages shown represent how many of the detected fixations from all the ones detected fell in that region. In the text, the fixation percentage presented is related only to the regions showing relevant content, that is, the percentage on the instructor over the sum of all fixation percentages over the other illustrated regions.

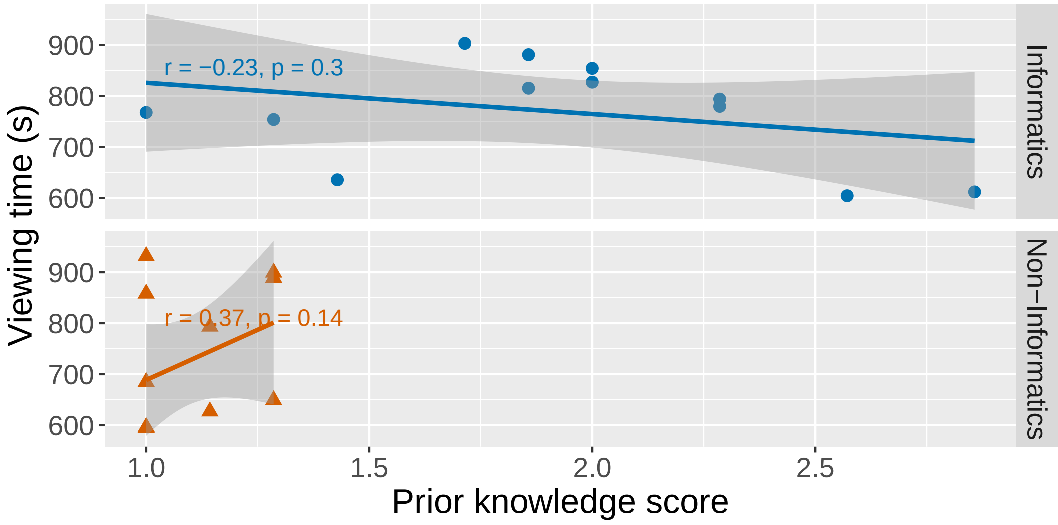


Figure 4.8 – Correlation between prior knowledge score [1-5], and viewing times for the different educational profiles. Regression lines are shown with a 95% confidence region.

Regarding the influence of prior knowledge on viewing session length (Figure 4.8), no significant correlations were found neither for Informatics subjects (Kendall's tau coefficient: $\tau = -0.32, p = 309$), nor for non-Informatics subjects ($\tau = 0.37, p = 138$). In addition, for Informatics profile, the observed correlation is negative. Hence, not only there is not enough evidence to validate hypothesis H_8, but also there is indication of a possible slight opposite effect.

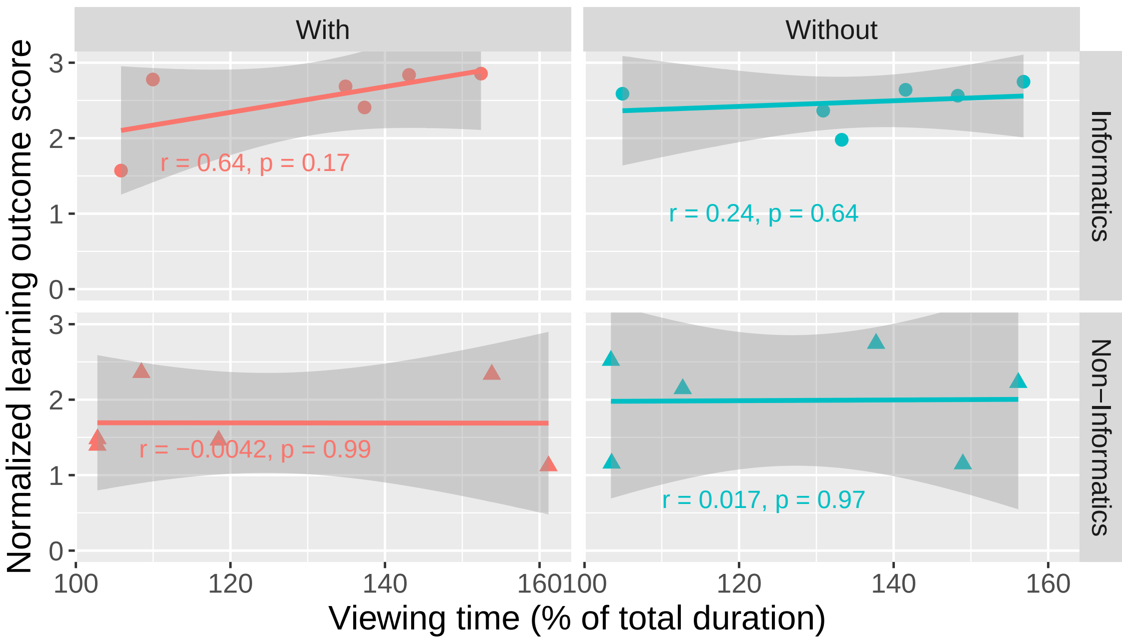


Figure 4.9 – Relationship between viewing time and learning outcome scores for each group of subjects ($n = 6$). Viewing times are given as the percentage of the total duration of the video lecture, that is, approximately 10 minutes. Regression lines are shown with a 95% confidence region.

Furthermore, the relationship between viewing time and learning performance was also analyzed (Figure 4.9). Learning outcomes do not significantly increase the longer the duration of the viewing session ($n = 24$). For the group of subjects from Informatics profile in *with instructor* condition ($n = 6$), a Kendall's tau correlation of 0.6 was observed between viewing times and total learning scores. However this correlation is not significant ($p = 0.136$). For lack of empirical evidence, hypothesis H_9 is not supported.

Both implications and limitations of the presented results are discussed in the next section. In chapter 5 comments on other possible analysis over the same data are also introduced.

— 5 —
Conclusion

The present study analyses two video lecture designs aiming at a better understanding of the process of viewing multimedia instruction, and as a result, to provide guidelines about which one of them is more effective for learning. Two contradicting hypothesis currently permeate the research on effectiveness in multimedia e-learning: the social-cue hypothesis, and the interference hypothesis. The first states a beneficial impact of instructor presence on both learning outcomes and learners' affective response. The latter enounces a negative effect of having a instructor visible, that such set of complex visual stimuli hinders learning. The results of an initial analysis of the data collected in our pilot experiment refute the social-cue hypothesis in the affective response sphere, and tend to support the interference hypothesis in terms of learning outcomes. Our results also validate the impact of instructor presence in learners' attention distribution, justifying the research on what are the effects of this presence on multimedia e-learning.

5.1 Discussion of results

Our results show that learners watching the video lecture without instructor visible tend to be more satisfied, have a better affective response, and perceive having learnt more. In contrast, other studies uncovered an affective benefit of the instructor face [**kizilcec2014showing**], and higher perceived learning in instructor present condition [**kizilcec2015instructor**]. We believe the subject area of the video lecture used in the experiment of each study may explain some of the differences in these results, due to different types of support materials needed for Sociology and Informatics classes. Learners do not like the fact that the instructor frame changes of position during the video, because it demands a lot of attention just to relocate the instructor position. In our experiment, the mobile instructor frame, combined with the lack of pointing to relevant content by the instructor might explain the negative affective response to instructor presence, as it imposes an extra not helpful element to be processed.

In terms of learning outcomes, the results here presented comply with some evidences in the field, but also contradict others. Instructor presence was found beneficial for recall on an easy Trigonometry topic [**wang2017instructor**], and for retention of spoken explanations [**colliot2018understanding**]. On the other hand, many studies found no significant differences in learning outcomes between instructor present and instructor absent video lecture designs (e.g., [**kizilcec2015instructor**, **van2017effects**, **wang2017instructor**]). Kizilcec et al. [**kizilcec2015instructor**] found in a large scale study ($n = 2,951$) that 35% of students prefer to watch video instructions without face of the narrator in order to avoid distraction. The present study did not find any significant differences in recognition, retention, and transfer scores. Nevertheless, differences on learning outcomes were still detected. Recognition scores were higher in *with instructor* condition for all learners, evidencing the advantages of nonverbal cues for this level of outcome. Retention was better for all learners in the *without instructor* condition, result leaning towards the interference hypothesis. Transfer scores and total scores differed in terms of most effective video lecture design for learners of different educational profiles, with Informatics performing better when instructor was present and non-Informatics when instructor was absent. We suppose these differences are due to the fact that, as non-Informatics learners know less about the content, the intrinsic cognitive load imposed to them by the task is higher, causing the image of the instructor to be an extraneous cognitive load. In addition, as the difference between mean learning scores were more important for the instructor absent condition, we consider that our results tend to support the interference hypothesis.

A curious result found is that retention scores were slightly lower than transfer results (see Table 4.1). This difference is not anticipated, as retention usually demands a lower comprehension of the content than transfer, as the latter requires from the learner the ability to reproduce the content learnt and not only be able to explain it, as in the case of retention. We estimate this result might indicate that our transfer question tested near transfer, and it may be interesting for future studies to also test far transfer. Due to this unexpected result, we believe the difference between retention and transfer scores merits further analysis in a subsequent study.

Regarding learners' attention distribution, our findings are similar to the others in the field. The presence of the instructor has a fair impact on learners viewing patterns, attracting, in average, 11.59% of learners detected fixations. Other studies found stronger impacts. For example, Kizilcec et al. detected a 41% portion of the time was spent looking at the face of the instructor [**kizilcec2014showing**]. T. van Gog et al. found that instructor video distracts learners from content, identifying that 30% of the time was spent looking at instructor face in detriment of relevant content [**van2017effects**]. We presume the lower percentage observed in the present study is due to differences in support material structure, i.e., the amount of text and images in the lecture slides. In addition, we uncovered differences in how instructor presence impacts attention distribution for different types of knowledge. Qualitatively, learners look less at the instructor video for procedural knowledge than for factual and conceptual knowledge. We suppose this phenomenon may be related to the amount of information to absorb for each type of knowledge. However, more research in this sense is necessary to create a deeper understanding of this effect. In summary, the relevant impact the presence of an overlaid instructor has on learners attention distribution – confirmed by the present study - justifies the research done in order to define the effects this presence has on learning effectiveness.

5.2 Limitations and research outlook

The present research conducted an initial analysis over a pilot experiment done in a limited, and rather short period of time. Therefore, the method and results here exposed present a few limitations. Firstly, we believe that with a more substantial and varied pool of subjects we would find a more important amount of statistically significant results. Other studies in the field present around 40 participants (e.g., 22 in [**kizilcec2014showing**], 36 in [**colliot2018understanding**], and 54 in [**van2017effects**]), and ours counted with 24 subjects. The participation of 2 non-native French speakers in the experiment might have an impact on the sample that could be further analyzed. In addition, our experimental protocol can be used to explore a broader set of video lecture designs. None of the two explored in this study, for example, presented any types of interaction with the content. Such lack of interaction might have influenced the non observance of the social-cue hypothesis, as the amount of nonverbal cues provided by the instructor may not be enough when the instructor stares straight at the camera. In this sense, some of the learners commented they felt a lack of signaling about what was the relevant content to be absorbed in each slide, mostly when the slide was heavy on content and the instructor was visible. Therefore, more research must be done to provide researchers with more insights on effects of interaction on learners behavior. This could be done, for example, by applying our experimental protocol to more video lecture designs and with a more substantial and varied pool of subjects.

The experimental protocol developed during the present research would also benefit from some modifications. The first one comes from a phenomenon detected when analyzing the collected data on perceived learning. Results show that, contrary to our hypotheses, a positive significant correlation between perceived learning and learning outcomes is present for non-Informatics subjects. We consider this might be an effect of the interpretation subjects made of the perceived learning question: "*Évaluez dans quelle mesure vous avez appris.*" (Assess how much you have learnt.). Based on our own re-interpretation of the question and on participants comments, we suppose subjects answered how many new things they have learnt, instead of the intended how much of the content have they learnt. Having based the perceived learning measure from other studies, the occurrence of such misinterpretation produces evidence for the importance of providing the original questions given to subjects in an experiment. In face of the subjects interpretation of the question, the correlation found for non-Informatics learners becomes understandable, as people from other fields usually have may face more completely new information from an Semantic Web video lecture than people from Informatics. So, recognizing this limitation, we suggest that, in future applications of the questionnaire, the perceived learning question should be rephrased to "*Évaluez dans quelle mesure vous avez appris le contenu du cours.*" (Assess how much of the lecture content you have learnt).

Another enriching modification to the protocol would be measuring the time subjects take to solve the learning assessment test. Time spent to perform a task can be used as a measure of cognitive load demanded by that task [**van2008instructional**]. Having the time subjects needed to solve the questions about each type of knowledge, for example, allows to compute the condition efficiency measure for factual, conceptual, and procedural knowledge. This would allow more insights about the differences in learning processes for each type of knowledge, and allow the researcher to provide more robust guidelines about effectiveness of video lecture designs for learning.

The interest of measuring learners cognitive load goes beyond post-measures of instructional condition efficiency. Imagine conclusive results have been produced about which video lectures designs are more effective for learning in each situation, based on the demand of learners cognitive load. Then, if one could measure viewers cognitive load in an on-line manner, an efficient adaptive learning environment would be possible. That is, learners cognitive load is detected on-the-go while watching the multimedia lecture, and then the pre-defined guidelines for the identified load or change of load are applied. Therefore, research on observational digital measures of cognitive load is essential. The detection of viewers cognitive load can be made through the analysis of observational data, such as pupil dilation, and facial expressions [**salojarvi2005inferring**]. So, as the next phase of this project, we are interested in evaluating the possible uses of our experimental equipment in measuring subjects cognitive load while watching the video lecture. We intend to use the observational data already collected (i.e., body posture, pupil dilation, and facial expressions) in this pilot study and cross it against the self-reported cognitive load. The measure of instruction condition efficiency (*cf.* [**paas1994measurement**, **van2008instructional**], explained in section 3.4) might also be explored in this context. Both measured self-reported mental effort (i.e., while watching the video lecture, and while solving the learning assessment test) can be crossed with the observational data and compared against learning performances, providing another element to determine which video lecture design is more effective for learning.

The present study is a first phase of bigger project that is to be continued by the student Laura Lassance (the document redactor) in her Brazilian home University. This extension of the here-presented research will be developed in partnership with the Polytechnic School of São Paulo (Brazil), serving as the four months Final Project of Laura Lassance. Furthermore, we would like to inform that all the collected data, and the initial analysis made over it will be made available as open data, in following the movement of Open Science.

— A —

Tree of video lecture's content

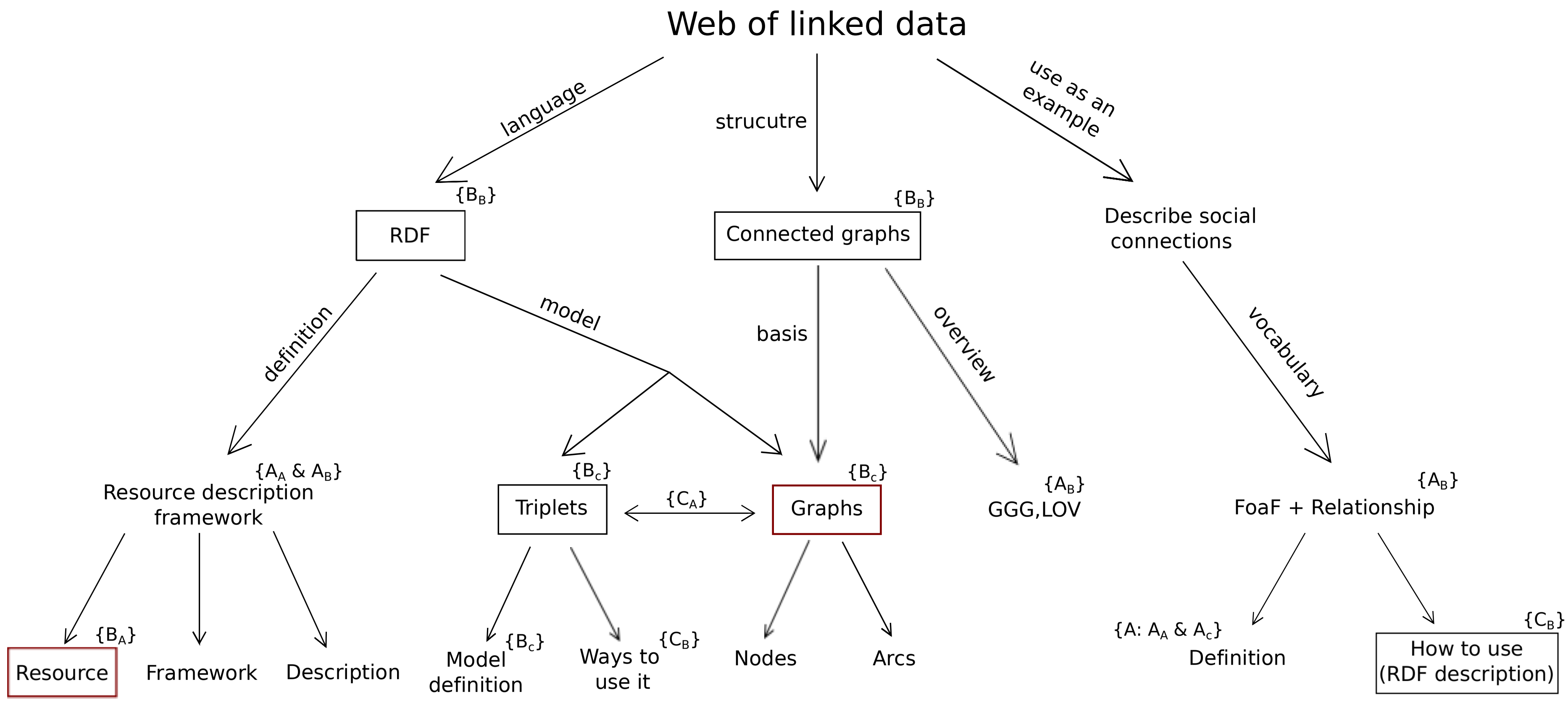


Figure A.1 – Tree organizing our understanding of the content presented in the video lecture used in this study about Web of Linked Data and Semantic Web. The main concepts are contained by a rectangle, and the essential ones have their rectangles indicated in red. The letters between curly brackets indicate the topic classification in terms of the type of knowledge the topic mostly represents. The classification into types of knowledge is based in the taxonomy presented in section 2.1

— B —
Consent form

Here are presented the consent information sheet and form participants needed to sign in order to participate in the experiment. It explains how data will be treated, and guarantees anonymity of the data. It also informs the participant that he / she can ask to stop the experiment at any moment, and that at any given time, even in the future, the participant can ask to have his data erased.

FORMULAIRE D'INFORMATION ET DE CONSENTEMENT

I. INTRODUCTION

Vous avez été invité(e) à participer à une étude sur l'analyse de l'utilisation des vidéos tutorielles comme un outil pour l'apprentissage. N'hésitez pas à poser des questions de façon à être bien sûr(e) de comprendre la procédure de l'étude, les implications de votre participation, incluant les risques et bénéfices.

II. BUT DE L'ETUDE/DU PROJET

Le but de cette étude est l'observation des trajectoires oculaires et des réactions physiologiques d'un étudiant d'un MOOC (Massive Open Online Course) dans le but de comprendre son interaction avec la vidéo et, aussi, l'effectivité de la vidéo par rapport à l'apprentissage.

III. PARTICIPANTS A L'ETUDE

Nous faisons appel à des participants volontaires pour cette recherche. Ce formulaire de consentement a pour but de vous informer sur cette étude, ses buts et ce que votre participation impliquerait. Merci de poser librement toutes vos questions et de ne pas signer si vous ne comprenez pas certains aspects de l'étude et de ses objectifs. La participation est totalement libre et volontaire. **Vous pouvez l'interrompre à tout moment sans encourir aucun préjudice ni conséquence.**

IV. PROCEDURE

L'expérience consiste à regarder une leçon en vidéo et après répondre une dizaine de questions par rapport aux ressentis pendant le visionnage de la leçon et la compréhension du son contenu. La durée de cette expérience est égale à la durée de la vidéo (10 minutes) plus le temps que le participant prend pour répondre aux questions (estimée entre 25 et 35 minutes).

V. RISQUES ET INCONVENIENTS

Il n'existe pas de risque prévisible ou attendu dans cette recherche. Elle nécessite simplement d'être disponible pendant 35-45 minutes pour y participer.

VI. BENEFICES

En participant à cette étude, vous contribuez de façon importante au développement de technologies futures. Les domaines d'applications incluent des systèmes d'accompagnements et de suivis pour les personnes âgées, des systèmes améliorant le dialogue avec des personnes présentant des difficultés d'expression (paralysie, autisme..), des tutoriels en support à l'acquisition de nouvelles compétences dans divers domaines etc.

VII. CONFIDENTIALITE

Les données, une fois recueillies, seront enregistrées de façon anonyme dans une base de données puis elles seront traitées pendant la phase d'analyse des données. Ils seront gardées pour la durée de cinq ans. Les informations seront utilisées par le équipe de recherches du projet : l'équipe Pervasive Interaction (INRIA). Les résultats de ces analyses et les vidéos et images provenant des enregistrements du setup expérimental seront anonymisées afin de retirer toute information personnelle permettant de vous identifier avant de les intégrer dans des rapports internes du projet ou plus tard dans des publications scientifiques. Il ne sera pas possible de vous identifier dans aucun des rapports de l'étude, conformément à la législation en vigueur. Les vidéos et images ne seront pas diffusées sur Internet.

Les résultats de cette étude pourront être publiés dans des revues scientifiques ou diffusés lors de conférences. L'anonymat et la confidentialité des données recueillies est garantie par l'ensemble des partenaires de ce projet. Conformément aux dispositions de la CNIL (loi relative à l'informatique, aux fichiers et aux libertés), vous disposez d'un droit d'accès et de rectification.

L'autorisation d'utiliser vos données reste volontaire, elle reste valide jusqu'à la fin de l'étude à moins que vous ne décidiez de la retirer avant. Si vous décidiez de retirer votre consentement à participer, merci de contacter l'investigateur en charge de l'étude.

A partir du moment où vous vous retirez de cette étude, vos données ne seront utilisées dans aucune des phases ultérieures dudit projet. Cependant les documents qui auraient déjà été publiés ne pourront pas être modifiés.

VIII. PERSONNE CONTACT

Pour plus d'informations sur cette étude, vos droits en tant que participant, si vous n'étiez pas satisfait ou si vous aviez des questions concernant cette recherche, merci de bien vouloir contacter la personne suivante :

DESSUS Philippe

Équipe Pervasive Interaction / Inria

655, avenue de l'Europe

Montbonnot

38334 Saint Ismier Cedex

philippe.dessus@univ-grenoble-alpes.fr

CONSENTEMENT

Votre participation à cette étude n'est possible que si vous signez librement ce formulaire de consentement et autorisez l'utilisation des données recueillies. Si vous ne le souhaitez pas, merci de ne pas signer et de ne pas participer.

Je soussigné(e), Mme, M[1] (*nom, prénom)*..

accepte librement de participer à la recherche intitulée « Analyse des signaux de communication dans les cours en-ligne » :

- j'ai pris connaissance de la note d'information m'expliquant l'objectif de cette recherche, la façon dont elle va être réalisée et les implications de ma participation,
- je conserverai un exemplaire de la note d'information et du consentement,
- j'ai reçu des réponses adaptées à toutes mes questions,
- j'ai compris que ma participation est libre et que je pourrai l'interrompre à tout moment, sans subir de préjudice ni conséquences. J'indiquerai alors à l'investigateur, si je souhaite ou non que les données recueillies, jusqu'au moment de ma décision, soient utilisées,
- j'ai bien été informé(e) que ma participation à cette recherche durera entre 35 et 45 minutes,
- mon consentement ne décharge en rien l'équipe de ce projet de l'ensemble de ses responsabilités et je conserve tous mes droits garantis par la loi,
- j'ai été informé(e) que les informations et les enregistrements me concernant et recueillies au cours de cette étude resteront confidentielles et ne pourront être consultées que par l'équipe de recherche.

Signature de la personne participant à la recherche	**Signature de l'investigateur**
Nom Prénom :	Nom Prénom :
Date : Signature :	Date : Signature :

[1] *rayer les mentions inutiles*

CONSENTEMENT

Votre participation à cette étude n'est possible que si vous signez librement ce formulaire de consentement et autorisez l'utilisation des données recueillies. Si vous ne le souhaitez pas, merci de ne pas signer et de ne pas participer.

Je soussigné(e), Mme, M[2] (*nom, prénom)*..

accepte librement de participer à la recherche intitulée « Analyse des signaux de communication dans les cours en-ligne » :

- j'ai pris connaissance de la note d'information m'expliquant l'objectif de cette recherche, la façon dont elle va être réalisée et les implications de ma participation,
- je conserverai un exemplaire de la note d'information et du consentement,
- j'ai reçu des réponses adaptées à toutes mes questions,
- j'ai compris que ma participation est libre et que je pourrai l'interrompre à tout moment, sans subir de préjudice ni conséquences. J'indiquerai alors à l'investigateur, si je souhaite ou non que les données recueillies, jusqu'au moment de ma décision, soient utilisées,
- j'ai bien été informé(e) que ma participation à cette recherche durera entre 35 et 45 minutes,
- mon consentement ne décharge en rien l'équipe de ce projet de l'ensemble de ses responsabilités et je conserve tous mes droits garantis par la loi,
- j'ai été informé(e) que les informations et les enregistrements me concernant et recueillies au cours de cette étude resteront confidentielles et ne pourront être consultées que par l'équipe de recherche.

Signature de la personne participant à la recherche		**Signature de l'investigateur**	
Nom Prénom :		Nom Prénom :	
Date :	Signature :	Date :	Signature :

---

2 *rayer les mentions inutiles*

— C —
Profiling pre-questionnaire

The pre-questionnaire given to participants before they were asked to watch the video lecture is presented here. It contains three profiling questions about participants educational background, educational level, and prior knowledge in the content of the video lecture to be watched.

PRE-QUESTIONNAIRE

1. **Quel est le domaine principal de vos études ?**

- Arts et Sciences Humaines
- Economie et Gestion
- Sciences Sociales
- Santé et Médecine
- Sciences de la Nature et Mathématiques
- Ingénierie et Technologie
- Autre

2. **Quel est votre niveau d'études (dernier terminé ou courant) ?**

- Avant le Baccalauréat
- Baccalauréat
- Licence (Bac+1 à Bac+3)
- Master (Bac+4 à Bac+5)
- Doctorat
- Après le Doctorat

3. **Avez-vous des connaissances de base en Informatique (programmation, bases de données, concepts de base, etc.) ?**

☐ **Oui** ☐ **Non**

Si oui, évaluez votre niveau de connaissance des thèmes sur le Web Sémantique listés ci-dessous.

	1 Aucune connaissance	2	3	4	5 Expert
Web Sémantique	○	○	○	○	○
URI	○	○	○	○	○
RDF	○	○	○	○	○
Graphes	○	○	○	○	○
Triplets	○	○	○	○	○
Vocabulaire FoaF	○	○	○	○	○
Vocabulaire Relationship	○	○	○	○	○

— D —

Engagement, satisfaction, and perceptions questionnaire

The affective post-questionnaire given to participants was made in two versions: one for subjects in the *without instruction* condition, and another one for subjects in the *with instruction* condition. Two versions were produced due to the question about the affective response to instructor presence. In the *without instruction* version, the presence is defined as the voice ("*presence par son discours*"), while in the *with instruction* version the presence is described as image and voice ("*presence par son image et son discours*"). Both questionnaires are presented here, in the order *without instruction* version, then *with instruction* version.

POST-QUESTIONNAIRE

1. **Lisez attentivement chaque phrase et répondez, sur l'échelle située en face, en entourant un nombre correspondant le mieux à ce que vous pensez.**

Quand j'ai visionné la leçon en vidéo…

	1 Pas du tout d'accord	2 Très peu d'accord	3 Un peu d'accord	4 Moyennement d'accord	5 Assez d'accord	6 Fortement d'accord	7 Tout à fait d'accord
J'étais totalement absorbé(e) par ce que je faisais	○	○	○	○	○	○	○
J'ai eu le sentiment de vivre un moment enthousiasmant	○	○	○	○	○	○	○
Je n'ai pas vu le temps passer	○	○	○	○	○	○	○
Cette activité m'a procuré beaucoup de bien-être	○	○	○	○	○	○	○
J'étais profondément concentré(e) sur ce que je faisais	○	○	○	○	○	○	○
Quand j'évoque cette activité, je ressens une émotion que j'ai envie de partager	○	○	○	○	○	○	○

2. **Évaluez dans quelle mesure vous avez appris.**

N'ai rien appris — Ai appris beaucoup

1	2	3	4	5	6	7	8	9

3. **Dans quelle mesure êtes-vous satisfait(e) de l'apprentissage avec cette vidéo ?**

Extrêmement insatisfait(e) — Neutre — Extrêmement satisfait(e)

1	2	3	4	5	6	7	8	9

4. **Au regard de la qualité de la vidéo, évaluez dans quelle mesure vous êtes d'accord avec la déclaration suivante :**

J'ai envie de visionner les prochaines leçons (vidéos) de ce cours.

Pour répondre à cette question, supposer que vous êtes intéressé au sujet.

Pas du tout d'accord — Indifférent — Tout à fait d'accord

1	2	3	4	5	6	7	8	9

5. **A combien estimez-vous l'effort investi dans le visionnage et la compréhension de la leçon ?**

Effort mental très, très faible								Effort mental très, très élevé
1	2	3	4	5	6	7	8	9

6. **Décrivez le rôle de la présence de l'enseignant (par son discours) sur votre compréhension de la vidéo :**

   a. Cette présence était-elle **aidante ?**

Pas du tout aidante									Très aidante
1	2	3	4	5	6	7	8	9	10

   b. Cette présence était-elle **utile ?**

Pas du tout utile									Très utile
1	2	3	4	5	6	7	8	9	10

7. **Décrivez le rôle de la présence de l'enseignant (par son discours) sur votre ressenti à propos de la vidéo :**

   a. Cette présence était-elle **frustrante ?**

Très frustrante									Pas du tout frustrante
1	2	3	4	5	6	7	8	9	10

   b. Cette présence était-elle **gênante ?**

Très gênante									Pas du tout gênante
1	2	3	4	5	6	7	8	9	10

   c. Cette présence était-elle **confuse ?**

Très confuse									Pas du tout confuse
1	2	3	4	5	6	7	8	9	10

POST-QUESTIONNAIRE

1. **Lisez attentivement chaque phrase et répondez, sur l'échelle située en face, en entourant un nombre correspondant le mieux à ce que vous pensez.**

Quand j'ai visionné la leçon en vidéo…

	1 Pas du tout d'accord	2 Très peu d'accord	3 Un peu d'accord	4 Moyennement d'accord	5 Assez d'accord	6 Fortement d'accord	7 Tout à fait d'accord
J'étais totalement absorbé(e) par ce que je faisais	○	○	○	○	○	○	○
J'ai eu le sentiment de vivre un moment enthousiasmant	○	○	○	○	○	○	○
Je n'ai pas vu le temps passer	○	○	○	○	○	○	○
Cette activité m'a procuré beaucoup de bien-être	○	○	○	○	○	○	○
J'étais profondément concentré(e) sur ce que je faisais	○	○	○	○	○	○	○
Quand j'évoque cette activité, je ressens une émotion que j'ai envie de partager	○	○	○	○	○	○	○

2. **Évaluez dans quelle mesure vous avez appris ?**

N'ai rien appris								Ai appris beaucoup
1	2	3	4	5	6	7	8	9

3. **Dans quelle mesure êtes-vous satisfait(e) de l'apprentissage avec cette vidéo ?**

Extrêmement insatisfait(e)				Neutre				Extrêmement satisfait(e)
1	2	3	4	5	6	7	8	9

4. **Au regard de la qualité de la vidéo, évaluez dans quelle mesure vous êtes d'accord avec la déclaration suivante :**

J'ai envie de visionner les prochaines leçons (vidéos) de ce cours.

Pour répondre à cette question, supposer que vous êtes intéressé au sujet.

Pas du tout d'accord				Indifférent				Tout à fait d'accord
1	2	3	4	5	6	7	8	9

5. **A combien estimez-vous l'effort investi dans le visionnage et la compréhension de la leçon ?**

Effort mental très, très faible								Effort mental très, très élevé
1	2	3	4	5	6	7	8	9

6. **Décrivez le rôle de la présence de l'enseignant (par son image et son discours) sur votre compréhension de la vidéo ?**

   a. Cette présence était-elle **aidante ?**

Pas du tout aidante									Très aidante
1	2	3	4	5	6	7	8	9	10

   b. Cette présence était-elle **utile ?**

Pas du tout utile									Très utile
1	2	3	4	5	6	7	8	9	10

7. **Décrivez le rôle de la présence de l'enseignant (par son image et son discours) sur votre ressenti à propos de la vidéo ?**

   a. Cette présence était-elle **frustrante ?**

Très frustrante									Pas du tout frustrante
1	2	3	4	5	6	7	8	9	10

   b. Cette présence était-elle **gênante ?**

Très gênante									Pas du tout gênante
1	2	3	4	5	6	7	8	9	10

   c. Cette présence était-elle **confuse ?**

Très confuse									Pas du tout confuse
1	2	3	4	5	6	7	8	9	10

— E —

Learning assessment test

Here is reproduced the learning assessment test given to participants after they had finished watching the video lecture. Recognition is tested in questions 1 and 2, retention in question 3 to 9, and transfer in question 10.

ANALYSE DE L'APPRENTISSAGE

1. **Cochez dans la liste ci-dessous les sujets qui ont été traités dans la leçon que vous venez de visionner.** *(Plusieurs options possibles)*

☐ La composition d'un triplet
☐ Les modèles RDF et leurs principes
☐ La notion d'ontologie
☐ L'orientation des arcs d'un graphe RDF
☐ La définition et le structure du Web Sémantique
☐ La caractéristique locale et isolée, i.e., non connectée des graphes RDF

2. **Choisir dans la liste suivante les sujets qui n'ont pas été traités pendant la leçon que vous venez de visionner.** *(Plusieurs options possibles)*

☐ L'organisation et les sources du LOV
☐ La définition du sigle RDF
☐ Où utiliser et comment placer les URI et littéraux
☐ La pile des standards du Web de données W3C
☐ Comment représenter un énoncé sur une page en RDF
☐ L'utilisation du FoaF avec RDF
☐ Comment représenter des schemas RDF en utilisant RDFS
☐ Les syntaxes de sérialisation de données RDF

3. **L'exemple suivant représente quel vocabulaire ?** *(Une seule option)*

```
<http://ns.inria.fr/catherine.faron#me> a [redacted]:Person ;
  [redacted]:family_name "Faron Zucker" ; [redacted]:givenname "Catherine" ;
  [redacted]:homepage <http://www.i3s.unice.fr/~faron/> ;
  [redacted]:mbox <mailto:faron@unice.fr> ;
  [redacted]:knows <http://ns.inria.fr/fabien.gandon#me> ;
  [redacted]:knows <http://ns.inria.fr/olivier.corby#me> ;
  [redacted]:topic_interest <http://xmlns.com/foaf/spec/> .
```

○ RDF
○ FoaF
○ Relationship
○ LOV
○ GGG
○ URI
○ Je ne sais pas (JNSP)

4. Quelle est la structure de données du modèle RDF ? *(Une seule option)*

- O Un arbre XML
- O Une liste doublement chaînée
- O Un multi-graphe étiqueté orienté
- O Une matrice d'adjacente
- O Un triplet unique
- O Je ne sais pas (JNSP)

5. Que signifie le sigle RDF ?

R ……………………..……………………………
D ……………………..……………………………
F ……………………..……………………………

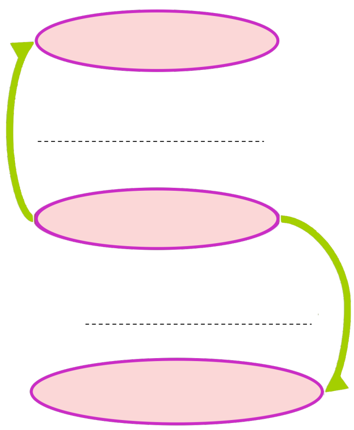
6. Quels sont les symboles utilisés dans des graphes RDF pour représenter des URIs et littéraux, respectivement ? *(Une seule option)*

- O Un **rectangle** représente les URIs et une **ellipse** les littéraux
- O Un **cercle** représente les URIs et un **rectangle** les littéraux
- O Une **ellipse** représente les URIs et un **rectangle** les littéraux
- O Un **rectangle** représente les URIs et un **losange** les littéraux
- O Un **losange** représente les URIs et un **rectangle** les littéraux
- O Je ne sais pas (JNSP)

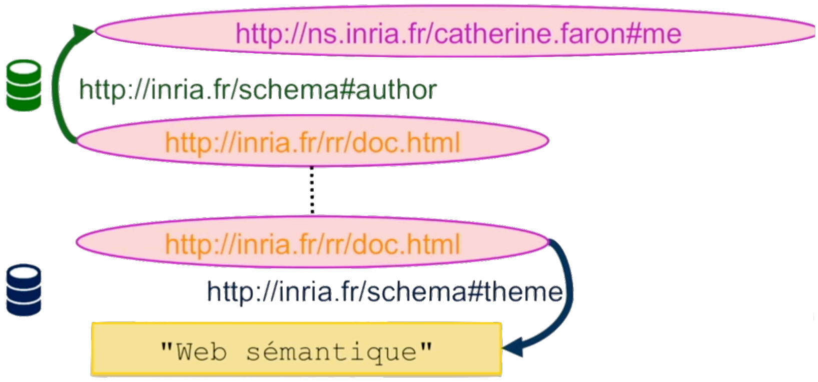
7. Comment modéliser en RDF l'énoncé suivant ? Remplir les espaces avec les données nécessaires, à la fois dans les triplets et dans la figure.
Utiliser seulement un des personnes sur le graphe.

"DOC.HTML A POUR AUTEURS FABIEN, CATHERINE ET OLIVIER ET PRÉSENTE LE WEB SÉMANTIQUE."

(**doc.html** , ,)
(, ,)
(, ,)
(, ,)



8. **Que signifient les icônes à gauche et le lien pointillé ?**



____

____

____

____

____

9. **Quelle est la différence entre une URI et un littéral ?**

____

____

____

____

10. Considerer l'énoncé suivant :

> "Je m'appelle Paul et j'ai 6 ans. Ma mère, Morgane, travaille avec le grandpère de Lucie (mon ennemie). Le grandpère de Lucie s'appelle Didier. Il a une page personnelle (didier.fr) dont le titre est Didier's page."

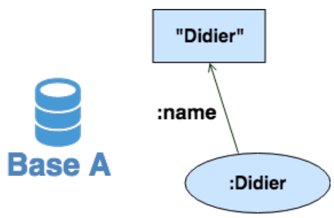
Les ressources représentant et décrivant **Didier (son nom et sa page personnelle)** sont stockées dans une base de données, que l'on appellera **base A**. Par contre, toutes les autres ressources sont stockées dans une autre base de données, la **base B**. Une des ressources peut être partagée entre les bases de données.

Construire un graphe RDF qui représente cet énoncé en considérant la division physique des ressources. Représenter les littéraux entre guillemets ("") et les URI après deux-points (:). Vous pouvez **utiliser les vocabulaires** présentés pendant la leçon et ci-dessous. Ces termes peuvent être utilisés plus d'une fois ou, aussi, aucune fois.

INFORMATIONS EXTRA

Contenu du ***FoaF*** et du ***Relationship***

:name	:title	:familyName	:knows	:age	:homepage	:topic_interest
:closeFriendOf	:enemyOf	:friendOf	:grandchildOf	:parentOf	:worksWith	



11. A combien estimez-vous l'effort mental investi dans la résolution de ces questions ?

a. **Questions 1 et 2**

Effort mental très, très faible								Effort mental très, très élevé
1	2	3	4	5	6	7	8	9

b. **Questions de 3 à 9**

Effort mental très, très faible								Effort mental très, très élevé
1	2	3	4	5	6	7	8	9

c. **Questions 10**

Effort mental très, très faible								Effort mental très, très élevé
1	2	3	4	5	6	7	8	9

COMMENTAIRES

— F —

In detail: taxonomy of video lecture designs

The video lecture designs used in the present study were base on Crook and Schofield's taxonomy. Figure 2.3 presents the formats identified and their classification. Here we present the complete explanation of the defined categories, as described in the original text (pages 59, 60):

A1 Voice over slides: A sequence of slides narrated by a hidden voice.

A2 Voice over screencast: A record of continuous screen recording (as opposed to discrete and static slides) is narrated by a hidden voice.

A3 Writing over slides: Narrated slides include superimposed the narrator's writing. Graphic annotation is added to one or more static images, implicitly by the speaker.

A4 Kahn whiteboard: Narrated whiteboards includes manual acts of superimposed writing. This is similar to A3, except that speaker's hand is made visible as they perform the annotation, thereby conveying a stronger sense of agency.

B1 Fixed frame outside: Video narrator in a window fixed adjacent to a slide sequence. The first of four formats that explore picture-in-picture presence of the lecturer. These may each vary in size but are generally small, typically occupying 20% of screen space.

B2 Mobile frame outside: Video narrator in a window in various positions adjacent to the sequence of background presentation activity.

B3 Fixed but overlapping: Video narrator at fixed position but overlapping the background sequence rather than being a *framed* picture in picture

B4 Mobile frame and overlapping: Video narrator is now framed, but present at varying positions in the background sequence.

C1 Presence in split screen: Video narrator and slide sequence are presented simultaneously and in adjacent frames.

C2 Presence in picture: Video narrator is visually integrated with slide images as if standing in front of a display surface.

C3 Presence overlapped by content: Symbolic material is superimposed on a video narrator.

D1 Presence active on whiteboard: Narrator moves in front of content and acts upon it but visual presence overlaps a full-screen presentation surface.

D2 Presence in lecture: Direct recording of narrator in traditional lecture context. The continuity of speaker and display surface is broken, conveying an in-room sense of the two.

D3 Presence in full screen: Close up on a solitary narrator in local "domestic" or topic-relevant context.

E1 Presence in interview: recorded interview.

E2 Presence in discourse: recorded conversation. This and E1 correspond to more traditional "talking heads" formats common in broadcast expositions.

Category A is characterized by the narrator being only a hidden voice. Its subdivisions have progressively more interaction of the narrator with the content displayed.