Can adapting teaching material to the learning style increase the learning performance in statistical subjects?

Damiano Gasparotto

Submitted in partial fulfilment of the requirements of the degree

of

MSc Information Management and Digital Business – Big Data in Business

Henley Business School

September, 2022

Abstract

The introduction of technological tools into education and training processes is commonly known as e-learning. Among the many advantages it has brought is flexibility in time, space, and mode, that is, the ability to take advantage of lessons from anywhere, at a time when it is most convenient, and in the way one prefers. In general, e-learning has facilitated greater personalization of learning, but this is still incomplete when considering the personalization of teaching materials and methods. In pedagogy, it is well known that there are differences in learning styles among individuals. The "one size fits all" problem occurs when individuals with different learning styles are provided with the same learning materials. Studies on the subject have shown that adaptive learning generally has a positive impact on learning performance. These studies, however, were conducted by adopting a learning style model that was designed for computer science students. This study aimed to extend learning performance assessment to other areas such as teaching statistics subjects. To carry out this study, Honey and Mumford's learning styles theory was chosen, in which four different styles are identified: Activist, Reflector, Pragmatic, Theoretical. The results obtained in this study are in line with the results obtained in the literature and demonstrate a positive effect of adaptive e-learning on learners. This paper also tried to contribute with the literature regarding the improving of learning styles detection but without obtaining relevant results.

Acknowledgements

I would like to express my appreciation to Dr Markos Kyritsis, who followed me through this research project, giving me meaningful advice and guidance in order to complete my studies to the best of my ability.

I would also like to thank all the participants in this research project who have shown me their support by giving me their time.

I would also like to thank the fellow students I met during this experience at the University of Reading, who helped to make it special.

Finally, the most heartfelt thanks I dedicate to my family and my fiancée who have always supported my choices, supported me in difficult times and shared the greatest joys with me.

Table of Contents

Introduction **1**
Motivations and background 1
Research Questions 2
Objectives and expected contribution 2
Research methods and project plan 3
Structure of dissertation 5
Literature Review **6**
Introduction 6
Definition of e-learning 6
Benefits and limitations of e-learning 7
Learning styles 9
VAK model 9
Kolb's model 10
Honey and Mumford model 10
Felder-Silverman model 12
Adaptive e-learning and previous works 13
Gap in literature and research questions 14
Methodology and Research design **16**
Introduction 16
Research design 16
Literature review 17
Learning styles detection 17
Quantification of Prior Knowledge 18
Research subjects 19
Sampling 20
Lessons and Learning materials 20
Data analysis 21
Ethical implication 22
Finding and Analysis **23**
Participants description 23
Learning style detection 28

Performance analysis 32
Analysis of relationship between learning styles and interactions 35
Discussion **37**
Limitations and suggestion for Future Research 38
Conclusion **39**
References **40**
Appendices **46**
Appendix I – Honey and Mumford questionnaire and demographic questions 46
Appendix II – Prior Knowledge Test 50
Appendix III – Final Test 52
Appendix IV – Lectures preview 55

List of Figures

Figure 1 – Honey and Mumford cycle 12
Figure 2 – Distribution of language chosen by participants 24
Figure 3 – Distribution of participants gender 24
Figure 4 – Distribution of participants educational level 25
Figure 5 – Distribution of participants employment 25
Figure 6 – Distribution of participants educational background 26
Figure 7 – Distribution of participants age 26
Figure 8 – Distribution of prior knowledge test score 27
Figure 9 – Cluster 1: differences in learning styles portions 29
Figure 10 – Cluster 2: differences in learning styles portions 30
Figure 11 – Cluster 3: differences in learning styles portions 31
Figure 12 – Cluster 4: differences in learning styles portions 31
Figure 13 – Distribution of results in final tests (“Mark”), divided by sample 33
Figure 14 – Distribution of results in final test (“Mark”) 34
Figure 15 – Distribution of final test results (“Mark”), divided by sample 35
Figure 16 – Activist Lesson preview 55

Figure 17 – Pragmatist Lesson preview 56
Figure 18 – Theorist Lesson preview 56
Figure 19 – Reflector Lesson preview 57

List of Tables

Table 1 – Felder-Silverman Model dimensions 12
Table 2 – Example of how questionnaire answers have been converted 18
Table 3 – Summary of Structure of the Knowledge Dimension of the Revised Taxonomy ... 18
Table 4 – Learning tools and objectives per each learning style 21
Table 5 – Summary results of prior knowledge test 27
Table 6 – Validation measures for cluster analysis with k = 4 28
Table 7 – Percentage of learning styles preferences in each cluster 32
Table 8 – Summary of final test results, divided by samples 33
Table 9 – Results of parametric assumptions check 36
Table 10 - 80-items Honey and Mumford Questionnaire 46

Introduction

Motivations and Background

The importance of education and training is well known and increasingly recognised both in the business world and in general by the population. An increasing number of people are completing tertiary education (Eurostat, 2022) and a growing proportion of companies are investing in staff training, thanks to studies that have shown the positive impact of employee training on company performance (Elnaga & Imran, 2013). In this regard, a current discussion focuses on how to improve the training services offered.

As in all sectors, technological development has had a strong influence leading to the evolution of traditional training systems. The introduction of e-learning, defined by Wang et al. (2010) as "use of computer network technology, primarily over or through the Internet, to deliver information and instructions to individuals", emphasised the personalisation perspective of the educational experience. Even before the advent of e-learning, however, personalisation was one of the recurring themes in education (Bartolomè et al., 2018).

The need to make the learning and teaching experience more flexible and tailored to individual needs has been partially met by the adoption of tools such as computers, the Internet and video recordings. Among the many advantages these have brought, it is possible to recognise temporal flexibility, which means being able to use educational materials as deferred recorded video lectures; spatial flexibility, namely the possibility of accessing lectures from any location; and modal flexibility, with the possibility of experiencing lectures in the way one prefers, repeating them several times, covering specific passages, skipping others.

However, flexibility is limited, as the "one size fits all" problem (Pagram & Pagram, 2006) persists with regard to the customisation of teaching materials and the methodology of teaching delivery. In response to this problem, the literature is oriented towards the development of new ways of obtaining knowledge and presenting different contents, with greater consideration of the learner's characteristics (Kulaglić et al., 2013). From the perspective of this type of personalisation, pedagogical studies on learning styles and differences in the learning process have a central importance (Barbe et al., 1979; Kolb, 1984; Honey & Mumford, 1982, Felder & Silverman, 1988).

The implementation of these studies in the educational and training environment falls under the branch called Adaptive E-Learning. The aim of the latter is "delivering the right content, to the right person, at the right time, in the most appropriate way-any time, any place, any path, any pace" (Shute & Towle, 2018). Adaptive e-learning is a topic that has only been partially studied and debated. The studies that have been conducted until now have yielded positive indications, which, nevertheless, deserve further investigation both in terms of the scope of application and in terms of the efficiency of the application. This research aims to fit into a partially developed context by attempting to extend the results obtained from previous studies to other disciplines and contexts.

Research Questions

The main research question of this research is:

- Can adapting teaching materials to Learning Styles improve learning performances in statistical subjects, during individual remote courses?

  To answer this question, we test the alternative hypothesis: H_1 = learning performance with adaptive e-learning > learning performance with normal e-learning

- Can the learning styles be identified using a knowledges entry test instead of traditional questionnaire?

  This question is to be answered by testing the existence of relationships between students' interactions with the entrance test and learning styles.

Objectives and expected contribution

This research aims to make an academic contribution to the existing literature on adaptive e-learning. In particular, as pointed out by Truong (2016), most studies use the learning styles theorised by Felder & Silverman (1988), while there are few studies on the application of Honey & Mumford's (1982) Learning Styles. We would like to study the results related to this learning styles perspective, as these were developed with an orientation related to problem-solving and decision-making, thus more akin to a business education. Whereas Felder & Silverman's model was developed considering computer science students and then extended to other categories. The aim of this project is to support studies demonstrating the benefits of adaptive e-learning and more specifically to examine the positive impacts of using the Honey & Mumford model with a view to enrich and diversify the literature on adaptive e-learning.

Data analysis lies at the intersection of statistics, computer science and business and is a discipline that is finding increasing interest in the world of work (Lovaglio et al., 2018). Considering the experiments carried out in adaptive e-learning, it is evident that no work had applied this tool to this subject area. Therefore, this work aims to explore whether the positive impact demonstrated in previous studies can also be extended to this discipline.

This study is also expected to make a practical contribution. Specifically, the exponential growth of the use of e-learning in the corporate sector (Brugess, 2021) and in the academic sector (Fraser, 2021) is evidenced. By testing the effectiveness of adaptive e-learning also with regard to statistical and data analysis disciplines, a new development of the offered educational services can be suggested, in order to increase the performance of the users with a positive impact in customer satisfaction and provider reputation.

Research methods and project plan

In order to achieve the defined objectives, it was decided to proceed with a practical experiment on a population of persons over the age of 18. In the first part of the experiment, participants were offered the Honey & Mumford 80-items questionnaire (Honey & Mumford, 1989) to identify their preferences in terms of learning style. They were then asked to answer a pre-test on their background knowledge in the topics covered. After collecting this data, the participants were divided into two independent samples and one sample was given a lesson that matched their learning style preferences (Sample A), the second sample was given a random lesson without considering the participant's learning style preferences (Sample B). Finally, all participants were asked to answer a final test on the topics proposed in the lessons to test the subjects' learning performance.

The project plan for this research basically consists of four main stages:

1. Problem Search and Methodology: In this stage, the problem was defined, and a gap was found in the literature, the research questions were formulated and the methodology to be adopted to carry out this research was defined. These steps took place in May and June 2022.
2. Tests development: In this phase, an entry test of knowledges was developed, based on the Structure of the Knowledge Dimension of the Revised Bloom's Taxonomy (Krathwohl, 2002). This test was developed as a tool for the formation of two independent samples that were not biased by prior knowledge. It was decided to adopt this structure in order to allow the replicability of the experiment. Secondly, it was decided to develop the tests following this structure, as has also been done in the literature (Raykova et al., 2011), because it allows the assessment of different aspects of knowledge. Furthermore, these tests were used to assess the existence of a relationship between Honey and Mumford's Learning Styles and the subjects' interaction with the test, as suggested by Costa et al. (2020). The other test that was developed was an eight-question final test to assess the participants' learning performance. This step was carried out in July 2022.
3. Lesson development: In this step, the teaching material to be administered to the participants was created. This material consisted of four presentations that covered a list of common themes inherent to data analysis but developed differently to suit the characteristics of the different Learning Styles, in accordance with the guidance provided by Bontchev & Vassileva (2011). This stage was carried out in August 2022.
4. Data Collection, Analysis and Discussion: This step consisted first of all of the participant recruitment, the completion of the questionnaire and the first test, and the analysis of the data collected to form the two independent samples. This was followed by the conduct of the lectures and the final test. The performed data analysis consisted in the first part of a descriptive analysis of the participants, the identification of learning styles through a cluster analysis and an interpretation of the test results for the formation of the samples. Secondly, the main research hypothesis was tested with an independent two-sample t-test, in order to check whether the final test results between the two samples have a significant difference. Furthermore, it was searched for the existence of

relationships between the learning styles and knowledge tests to answer the secondary research question. Finally, this step was concluded with the discussion of the results and the production of a report of the work performed. The tasks described here were carried out during August and September 2022.

Structure of dissertation

In order to answer the research question, the report begins with chapter 2 "Literature Review" in which there is a first part dedicated to e-learning and pedagogical studies on learning styles, and then continues with an overview of the adaptive e-learning studies in which these theories are applied. Chapter 3 "Methodology and Research Design" describes in detail the methodology used, how the materials and tests were developed, how the data were collected and analysed, how the two samples were formed and the precautions that were taken to prevent ethical risks. In chapter 4 "Analysis and Findings" the relevant characteristics of the participants are described, then a first analysis on the detection of learning styles is conducted and subsequently an analysis regarding the results of the final test is carried out. In chapter 5 "Discussion" these results are interpreted and the impact they can have in the academic and economic field is discussed, the limitations of this research and potential opportunities for future research are also highlighted. Finally, Chapter 6 "Conclusion" summarises the work done and the results it has brought.

Literature Review

Introduction

In this chapter, it is discussed what e-learning is, what its potential is, to arrive at a more specific understanding of adaptive e-learning. To illustrate what adaptive e-learning is and the state of the art there is a digression into pedagogy to explain what learning styles are and the different models that are considered. Finally, the gap in the literature will be analysed and an attempt will be made to fill it by investigating the research questions.

Definition of e-learning

All sectors over the recent decades are facing a technological revolution, therefore education has also implemented the use of technological devices in its programmes in order to improve outcomes. This use of technology can be defined as e-learning and numerous definitions are given in the literature (Hassenburg, 2009; Wang et al., 2010; Tayebinik & Puteh, 2012; Sangrà et al., 2012; Behera, 2013):

- "E-learning refers to the use of computer network technology, primarily over or through the internet, to deliver information and instructions to individuals" (Wang et al., 2010).
- "The term "e-learning" has emerged as a result of the integration of ICT in the education fields" (Tayebinik & Puteh, 2012).
- "E-learning is an approach to teaching and learning, representing all or part of the educational model applied, that is based on the use of electronic media and devices as tools for improving access to training, communication and interaction and that facilitates the adoption of new ways of understanding and developing learning" (Sangrà et al., 2012)
- "E-learning is a means of education that incorporates self-motivation, communication, efficiency, and technology. It is a flexible term used to describing a means of teaching through technology. E-learning refers to the use of Internet technologies to deliver a broad array of solutions that enhance knowledge and performance" (Behera, 2013).

Sangrà et al. (2012) emphasise that it is not possible to find a single, all-encompassing definition of e-learning, as learning needs change very rapidly and educational theories and tools have to be continuously adapted to these needs.

It is not possible to find a single definition, however, it is possible to recognise that e-learning consists of the collaboration of several disciplines, including computer science, communication technology and pedagogy.

Building on these definitions, Kumar Basak et al. (2018) analysed which perspectives should be considered when creating tools and materials for the application of e-learning. The most suitable perspectives are those identified by Clark (2007). These perspectives are:

- Cognitive Perspective: this considers how are the cognitive processes and how the brain works (Clark, 2007).
- Emotional Perspective: this perspective highlights the engagement and the motivation during the learning process (Clark, 2007).
- Behavioural Perspective: the factors involved in this perspective are the skills and behavioural outcomes, that can be connected with role-playing and application to on-the-job settings (Clark, 2007; Kumar Basak et al., 2018).
- Contextual Perspective: this includes the environmental and social aspects that can help better performances in learning process (Clark, 2007).

Benefits and limitations of e-learning

In the literature, several studies have identified the different advantages of e-learning. The first benefit that is widely recognised is easier access to a wide range of learning materials (Keller & Cernerud, 2002; Malik & Rana, 2020; Al Rawashdeh et al., 2021). In addition to this, there is also the possibility of offering content in a more appealing manner and having a greater variety of learning experiences (Clark, 2007; Al Rawashdeh et al., 2021).

Another aspect in which there is broad consensus concerns the possibility of better and more effective communication between students and teachers (Al Rawashdeh et al., 2021). Furthermore, e-learning helps to remove barriers that may hinder participation, such as the fear of public speaking in front of other students. E-learning motivates students to interact with others through discussion forums in which different points of view can be shared (Clark, 2007; Arkorful & Abaidoo, 2015). A further advantage relates to the economic aspect; for example, travel costs for students to physically reach the school can be removed, also allowing for wider participation (Clark, 2007; Radović-Marković, 2010).

These numerous advantages cover different aspects of the learner experience, but the main e-learning benefit is flexibility.Keller & Cernerud (2002), Arkorful & Abaidoo (2015), Clark (2007) and Radović-Marković (2010) argue that e-learning allows greater flexibility in terms of time and space as it is possible for the learner to use the learning materials in any place and with greater temporal flexibility. It should also be considered more flexible in terms of the mode of use, for example, some learners prefer to concentrate on certain parts of the course, while others prefer to review the entire course, nonetheless, it also allows for self-fruition, i.e., the possibility of studying at one's own pace and speed (Arkorful & Abaidoo, 2015).

These gains lead to a proven enhancement of performance both in terms of learning, as argued by Malik & Rana (2020), and in terms of satisfaction and engagement (Radović-Marković, 2010, Arkorful & Abaidoo, 2015).

A problem that not only affects e-learning but is carried over from traditional face-to-face methods is the "one size fits all" problem. As argued by Yang et al. (2019), with the introduction of new technologies and artificial intelligence, education should move towards a dimension of personalisation of methodologies. On this view, Gašević et al. (2016) highlighted how evaluation models should differ from subject to subject and must take into account students' experiences and backgrounds. Jones (2017) argued how the change of perspective in teaching law should be based on multiple intelligences. The latter was also partly echoed by Bennani et al. (2021), which underlined that even innovative methods such as the gamification of e-learning achieved partial results in cases where a differentiation of teaching approaches was not applied, suggesting that the development of teaching material, differentiated according to the student's learning style, seemed to be an approach to follow.

Thalmann (2014) argued that the most useful method to develop adaptive and personalised e-learning systems are learning styles, more than students' background or previous knowledge.

Truong (2016) pointed out that the integration of computer science and the psychological and pedagogical areas has gained significant interest in recent years. Specifically, a number of works have been produced on the application of learning styles to computer-based educational systems that have yielded positive results (Lin et al., 2013; Kurilovas et al., 2014).

Given the centrality that learning styles are acquiring in the e-learning sphere, it is only right to digress into the pedagogical sphere with a brief overview of existing models.

Learning styles

The term "learning styles" refers to the idea that individuals differ in terms of the education method that is most effective for them (Pashler et al., 2008). Proponents of learning styles theories argue that, in order to improve the quality of education, a diagnosis of learning styles and a subsequent adaptation of methodologies to the characteristics of students is necessary.

The most famous theories proposing learning styles models are VAK model (Barbe et al., 1979), Kolb's model (Kolb, 1984), Honey and Mumford model (Honey & Mumford, 1982), Felder-Silverman model (Felder & Silverman, 1988).

VAK Model

The Visual-Auditory-Kinesthetic (VAK) model uses the main sensory receivers to determine a person's learning style and focuses only on the external aspects of the learning process. Everyone uses all methods to receive information but typically one of these plays a dominant role. In light of this, it can be said that the division made is not clearly distinguished and there may be individuals who have a mixed preference or distinct preference according to task.

The learning styles defined by this theory are thus:

- Visual: those who prefer seeing, written or pictorial material, who often use written information, notes, and diagrams to learn.
- Auditory: those who prefer listening, who learn from oral instructions and write few and confused notes.

- Kinesthetics: those who need to touch and be physically involved, learn by experiencing movement.

Kolb's Model

This model includes both perception and information processing in the diagram. According to Kolb (1984) perception and information processing are a continuous cycle involving four stages: concrete experiences (feeling), on which we then reflect (watching) and from which we begin to understand by formulating abstract concepts (thinking). Finally, we apply what we have learnt to reality, arriving at the step of active experience (doing). Two dimensions can be recognised in this cycle, namely how information is perceived (feeling vs. thinking) and how information is processed (watching vs. doing).

The 4 different types of learning styles identified by this model are the combination of the different options in the two dimensions, perceiving and processing:

- Accommodators (doing and feeling): these are those who are inclined to learn from their real experiences.
- Divergers (feeling and watching): are those who are more inclined to use personal experiences and practical ideas to formulate their own theories.
- Assimilators (watching and thinking): are those who are able to extend their understanding of abstract concepts to develop new theories of their own.
- Convergers (doing and thinking): are those who better process abstract ideas to achieve concrete results.

Honey and Mumford model

This model is closely related to the Kolb model in that it is based on the same concepts but has been revisited from a problem-solving and decision-making perspective. The quality of this model is that it gives a more practical solution to the previously stated theory and also provides a questionnaire that allows the respondent to be identified in the proposed learning styles (Honey & Mumford, 1989).

The cycle steps described above have been transformed into:

1. Having an experience
2. Reviewing the experience
3. Concluding the experience
4. Planning the next steps

The four learning styles highlighted by this model are:

- Activists: These types of people prefer to learn by doing. Activists need to get their hands dirty with practical experiences. They have a receptive way of approaching learning and prefer to avoid theory.

- Reflectors: These individuals prefer to observe and contemplate reality from several points of view. They may refrain from intervening and prefer to observe by standing back. Before coming to a conclusion, they want to have examined the different possibilities to ensure that they reach an appropriate conclusion.

- Theorists: These individuals want to understand the assumptions behind the activities. They need models, ideas and truths with a specific purpose to participate in the learning process. They break down and reformulate ideas to achieve new hypotheses following a coherent method.

- Pragmatists: These learners apply learned concepts to their current reality. They prefer to avoid abstract concepts and ideas if they do not recognise a usefulness and an application to their direct experience. The pattern that is commonly used for learning is related to experimenting with new ideas and new methods applied to their actions.

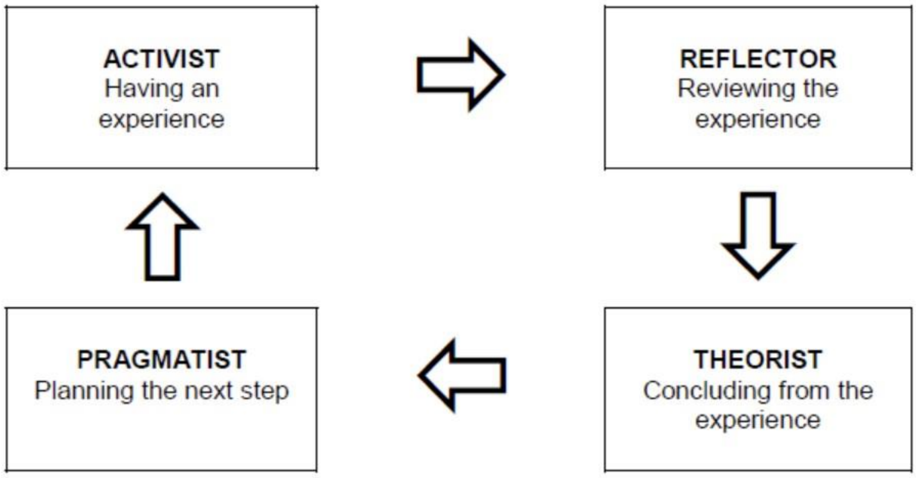

Figure 1 – Honey and Mumford cycle (Honey & Mumford, 1989)

Felder-Silverman model

This model was primarily designed to recognise the different types of learning styles among engineering students but can also have an application in other areas.

Initially, five dimensions were identified in which students can express a preference, but in 2002 a new edition of the model was proposed in which the dimensions were reduced to four.

The dimensions in which students can express their preference are collected in table 1.

Dimensions	Options
Preference in how to receive the information	Sensing vs Intuitive
Preference in how information is presented	Visual vs Verbal
Preference in information processing	Active vs Reflective
Preference in how organise and progress toward understanding information	Sequential vs Global

Table 1 – Felder-Silverman Model dimensions

From the combination of these preferences, learning styles are derived. The advantage of this model is that it covers more dimensions and thus allows for more nuances of the learning process. At the same time, this type of option leads to a higher granularity of the types of learning styles which makes it more difficult to customise the learning materials for each combination of preferences.

However, there are also studies that consider the existence of learning styles to be unreliable. In fact, Pashler et al. (2008) argued that there is no clear and conclusive evidence of the existence of learning styles and that this is mainly a claim to be able to sell new educational products. Reiner & Willingham (2010) and Rohrer & Pashler (2012) are also of the same opinion, but despite this, they agree that a one-size-fits-all solution for all learners is not the most effective method of lesson delivery.

Adaptive e-learning and previous works

Considering the need to develop new personalised education strategies, the application of machine learning and artificial intelligence models is an opportunity that has been partly addressed in the literature. This type of solution is called Adaptive E-learning and as argued by Kulaglić et al. (2013) scientific research in the field of e-learning is oriented towards the development of new ways of obtaining knowledge and presenting different content, with greater consideration of the expectations, motivation, learning styles, habits and needs of the learner.

The aim of adaptive e-learning, defined by Shute & Towle (2018) as "delivering the right content, to the right person, at the right time, in the most appropriate way-any time, any place, any path, any pace", can be adapted to all four perspectives proposed by Clark (2007) e Kumar Basak et al. (2018): Cognitive, Emotional, Behavioural and Contextual.

Focusing on the cognitive perspective, i.e. how are the cognitive processes and how the brain works, learning styles represent the declination of this idea as they focus precisely on the cognitive process and in addition are the most functional methods in the field of adaptive systems (Truong, 2016). For this reason, numerous works have been conducted on this topic in the literature.

Most of the work in the literature on adaptive e-learning is based on the Felder-Silverman Model (Truong, 2016), but according to Coffield et al. (2004), many theories have common and overlapping arguments, none of which, however, clearly prevails over the others in terms of reliability and performance.

As mentioned above, the Felder-Silverman model, although the most widely used, was developed taking into primary consideration the attitudes of computer science students (Felder & Silverman, 1988). Differently, the Honey-Mumford method was conceived from a more business-related perspective, considering ways of approaching problems and making decisions (Honey & Mumford, 1982). In addition, Saraswathy (2019) found the existence of a correlation between the Honey-Mumford model and learning approaches in mathematics. These reasons lead to further studies related to adaptive e-learning using the Honey and Mumford model.

The work carried out in the field of adaptive e-learning and learning styles can be divided into two categories: learning styles detection and performance evaluation. In the first category, studies differ in the type of input used such as personality questionnaires and context data (Vita, 2001; Paireekreng & Prexawanprasut, 2015), background knowledge, intelligent capability, cognitive traits (Germanakos et al., 2008). Other studies are distinguished by the different machine learning approaches used such as supervised learning (Paireekreng & Prexawanprasut, 2015) and unsupervised learning (El Aissaoui et al., 2019). Only Costa et al. (2020) used the Honey-Mumford model and tried to identify learning styles through interactions with the virtual learning environment during the lesson but did not obtain significant results.

With regard to studies that have investigated the performance of adaptive e-learning, papers are highlighted that have analysed the performance of engagement and interest in students, obtaining positive results in relation to the adoption of these systems (Cabada et al., 2011; Yang et al., 2013). On the other hand, other works focused on the study of learning performance (Essaid El Bachari & El Adnani, 2011; Latham et al. 2014; Hassan et al., 2021). All studies have shown positive increases in learning, but these studies were all conducted using the Felder-Silverman model and all applied adaptive systems to the teaching of computer science topics.

Gap in literature and research questions

As previously reported, there are studies that have evaluated the application of adaptive e-learning and have shown a positive effect on student learning performance. In all the proposed studies the Felder-Silverman model was applied, this may represent a potential gap, indeed there are no relevant studies that evaluate learning performance using other learning styles models, including the Honey-Mumford one.

Furthermore, the fact that adaptive e-learning was only conducted for teaching computer science is coherent with the choice of the Felder-Silvermann model, but at the same time limits the scope of the results. This research, exploiting the positive results that have been proposed by previous studies and also considering the correlation between the Honey-Mumford model and the learning of statistical-mathematical disciplines (Saraswathy, 2019), aims to assess whether the use of the Honey-Mumford model in the development of statistical lessons can improve learning performance.

Secondly, given the adoption of the Honey-Mumford model and the design of the researcher it is also possible to answer a secondary research question formulated on the findings of Costa et al. (2020). In particular, the aim is to understand whether there are relationships between duration and performance in the learner's interactions with the learning environment, by means of tests. This line of research fits in with studies that want to investigate learning styles detection.

In conclusion, therefore, it can be argued that the research questions aim to support the results obtained in previous works (Essaid El Bachari & El Adnani, 2011; Latham et al., 2014; Costa et al., 2020; Hassan et al., 2021), by reproducing them in a different field and with a different basic model. This makes it possible to assess whether the results obtained on adaptive e-learning are consistent and extendable.

Methodology and Research Design

Introduction

In this chapter, the methodology adopted in this work is described, starting with a brief description of the process, to arrive at a detailed explanation of all the steps performed and the discussion regarding possible ethical problems.

In order to answer the research questions, it was decided to perform a quantitative analysis of the problem, using the hypothetical-deductive approach proposed by Prodanov & De Freitas (2013), considered suitable because it was used by Costa et al. (2020) in research on Learning Styles applied to the Virtual Learning Environment. This method begins with the clear definition of the problem that facilitates the formulation of a theory and the identification of suitable tools that can contribute to the development of the research work. Subsequently, the behaviour of the formulated theory is observed through quantitative tests in which the performance of the proposed theory is evaluated. Eventually, the next stage consists of observing the results and formulating new hypotheses that should express what was noted in the previous stage.

In this case, the main research question aims to study the effectiveness of adaptive e-learning by assessing whether there is a significant difference in learning performance between those who receive material adapted to their learning style and those who receive material that differs from their learning style. Secondly, it aims to investigate whether there is a significant relationship between Learning Styles and learner interactions (duration and performance) with an initial knowledge test, developed following the Bloom's Taxonomy framework.

Research Design

In this work, in order to test the hypotheses formulated, it was decided to reproduce an empirical experiment whereby participants are asked to study a lesson and answer a test concerning the topics covered in the lesson. The experiment has begun with the determination of the participants' learning styles. For this purpose, participants were asked Honey and Mumford's (1989) 80-item questionnaire with some questions on demographics and their educational background. After completing the questionnaire, the participants were given an initial test of their prior knowledge about data analysis. Then after determining the participants' learning style and background knowledge, it was decided to divide them into two independent samples (sample A and sample B). Participants were engaged a second time where they were asked to study a short lesson on data analysis and complete a final test on the lesson studied. Sample A was provided with the lesson in an adaptive manner, so was assigned the material appropriate to the subject's characteristics, as argued by Shute & Towle (2018). In sample B, it was decided to randomly assign the material by simulating a learning context in which learning style preferences are not considered. The experiment has been concluded with a final test to assess how the participants' learning performance was.

Literature review

Initially, an analysis of the existing literature in the field of E-Learning, Learning Styles and Adaptive E-Learning was conducted in order to outline the state of the art and identify a gap in the literature to be studied. Google Scholar and the University of Reading website in the library section were used as search engines. Once the literature review was conducted, the methodology and tools to be used were determined.

Learning Style Detection

The learning style theory chosen for this Project is that formulated by Honey and Mumford (1982), which provides four different preferences in terms of style: Activist, Theorist, Reflector, Pragmatist. These preferences are detected through the use of a questionnaire (Honey & Mumford, 1989) consisting of 80 sentences, 20 for each learning style (Full questionnaire available in Appendix I). The participant is asked to select from the 80 sentences, those closest to the subject's way of acting and thinking.

The Questionnaire specifically does not exclusively identify a subject's learning style but identifies what the intensity of each learning style is in the respondent's preferences (Very Low, Low, Moderate, Strong and Very Strong). Both Honey & Mumford (1989) and later studies such as Alonso et al. (1997) and Costa et al. (2020) identified the subjects' learning style by assessing which one was predominant, without giving importance to intensity. On this basis, four new variables were created that identify the percentage of each learning style in the total number of responses given by the subject, as in the example shown in the following table 2

	Activist	**Theorist**	**Pragmatist**	**Reflector**	**Total**
Participant 1	9	2	6	6	23
	39%	9%	26%	26%	100%
Participant 2	10	12	8	11	41
	24%	29%	20%	27%	100%

Table 2 – Example of how questionnaire answers have been converted

Once these variables were developed, it was decided to use cluster analysis and specifically the k-means algorithm (with k = 4) in order to subdivide the participants on the basis of their predominant learning style. Microsoft Forms was used to collect the data and Microsoft Excel was used for the first data cleaning, finally, R studio was used for the cluster analysis.

Quantification of Prior Knowledge

In order to study what the learning outcomes were, it was necessary to quantify prior knowledge in the proposed lesson topics, balancing the samples that were created. To develop this knowledge test, it was important to adopt a framework that was defined by the literature, to make the experiment feasible and replicable. The choice therefore fell on the use of the Structure of the Knowledge Dimension of the Revised Taxonomy (Krathwohl, 2002) which breaks down knowledge into four sections described in table 3. The full test is in appendix II.

Sections of Knowledge Dimension	**Definition**
Factual Knowledge	The basic elements that students must know to be acquainted with a discipline or solve problems in it.

Conceptual Knowledge	The interrelationships among the basic elements within a larger structure that enable them to function together.
Procedural Knowledge	How to do something; methods of inquiry, and criteria for using skills, algorithms, techniques, and methods.
Metacognitive Knowledge	Knowledge of cognition in general as well as awareness and knowledge of one's own cognition.

Table 3 – Summary of Structure of the Knowledge Dimension of the Revised Taxonomy (Krathwohl, 2002)

Research subjects

This research focuses on the education and training of adult subjects who can be placed in university and working contexts. For this reason, the requirements for participation in the experiment were:

- Be of legal age (age > 18).
- Good English comprehension level, this in particular was necessary to minimise the bias generated by language difficulties.
- Participants must be in possession of a technological tool (Smartphones, PCs, Tablets...) and have access to an internet connection.

The search for participants was carried out among close contacts and acquaintances and only those who fulfilled the predefined conditions were involved.

Considering that previous studies such as Costa et al. (2020) and Hassan et al. (2022) had been carried out including 598 and 200 students respectively, it was not possible to replicate them for reasons of time and availability of participants, so it was decided to set a target number of 60 participants.

The search for participants did not lead to the achievement of a satisfactory number of participants (17 participants), so in order to increase the number of participants in the experiment, it was decided to overcome the condition of linguistic limitations by translating all the tests and learning materials into Italian, thus expanding the number of participants to 52 (of which 2 withdrew during the course of the experiment, reaching a final total of 50 participants).

Sampling

In this step, it was decided to operate an experiment with the creation of two independent samples. The choice of the two independent samples follows the choice made in other similar studies previously carried out (Hassan et al., 2022). Although the small number of participants suggested the adoption of a two paired samples technique, factors that could not be fully controlled such as the difficulty of the lesson topics, prior knowledge, and memorisation capacity over time would have affected the test results. Thus, it was preferred to divide the population into two independent samples using the stratified random sampling technique. In order to neutralise the effect of prior knowledge, the results of the initial knowledge test were used. The participants were divided into 6 subsets according to the grade obtained (e.g., subset 1 participants who obtained 1 point, subset 2 participants who obtained 2 points...). Then, each subset was divided into two, creating the two independent samples that had a homogeneous distribution of prior knowledge. The other condition that was placed on the division of the participants is that related to observations with doubtful or multiple learning styles (Honey & Mumford, 1989). It was therefore decided that participants with doubtful preferences would be placed in sample B, with random assignment of the learning materials, in order to avoid altering sample A with assignments that did not fully reflect the learning style.

Lessons and Learning Materials

This project focuses on individual study from a remote location, which is why learning activities involving student interaction and the typical dynamics of a classroom were excluded and all activities were carried out individually from a remote location. The learning materials produced are presentations covering the following topics:

- Introduction to data analysis
- Supervised and Unsupervised learning
- Linear Regression
- Variable Selection
- Subset Selection: backward and forward stepwise selection
- Shrinkage Methods: Ridge Regression and Lasso Regression

These topics have been developed differently, following the indications provided by Bontchev & Vassileva (2011), which have been summarised in table 4

LEARNING STYLE	LEARNING TOOLS
ACTIVIST	- Practice experiments - Problem solving - Examples - Single user games
REFLECTOR	- Classification - Comparisons - Examples
THEORIST	- Formalisations - Generalisation - Comparison - Case Studies
PRAGMATIST	- Exercises - Problem Solving - Intermediate tests - Case Studies

Table 4 – Learning tools and objectives per each learning style, identified by Bontchev & Vassileva (2011)

These lessons were developed following the topics covered by James et al. (2013). The presentations were elaborated with the use of the online software Prezi (Lectures preview in appendix IV).

Data analysis

After studying the proposed learning materials, the participants were asked to answer the final test. The score of the final test, consisting of 8 multiple choice questions (appendix III), ranged from 0 to 8 with 1 point for each correct answer and 0 for each incorrect answer. Once the data from the participants had been collected, it was decided to assess whether the mean score of sample A was higher than the mean score of sample B.

To do this, the first step was to test the parametric assumptions:

- Normality of the data: the normal distribution of the data was verified quantitatively with the Shapiro-Wilk test (Shapiro & Wilk, 1965), because it is the most powerful method especially with a small sample size (Razali et al., 2011).
- Homogeneity of variance: this assumption was verified using Levene's Test (Levene, 1960).
- Equal Sample Size: samples were created with the same number of elements.

Given the results of the previous tests it was then decided to perform an independent two samples one-tailed t-test. The alternative hypothesis to be tested is that the average score in the final test of the sample instructed with adaptive e-learning (sample A) is greater than the average score in the sample test where the learning style is not considered. For this reason, it is necessary to use a one-tailed t-test. Finally, the effect size was tested using Cohen's D, which is reliable and accurate because the assumptions of normality and homogeneity of variance have been tested previously (Cohen, 2013).

With the collected data we also tried to answer the secondary research question investigating the existence of a relationship between interactions in the initial knowledge test and learning styles (e.g., a reflective learner takes longer to answer the test, a theoretical learner performs better in procedural knowledge...). Assumptions were also checked, and then MANOVA tests were carried out using as dependent variables the interactions in the initial test and independent variables the learning styles defined beforehand.

Ethical implications

In this research work, the learning abilities and prior knowledge of a participating subject are studied. This data combined with the background data is configured as personal data by the GDPR (2016), and the subject's prior informed consent is required to use it. To increase the protection of participants and encourage their participation, it was decided to adopt a pseudonymisation measure. As this study required participation in two linked phases, participants were asked to enter a personal identification code, chosen by them. For the second phase where each participant was assigned a particular lesson, a table was created where each ID code was associated with the link to the correct learning materials and participants were asked to access the link associated with their code.

The Research Project and the materials used were approved by the Research Ethics Committee of the University of Reading.

Analysis and Findings

In this chapter, the results obtained from the data analysis are described and the main findings related to the proposed research questions are highlighted. In the first part, there is a description of the sample that participated in the study, followed by the definition of the participants' learning styles. Then, the results of the test performed on the main research question are demonstrated and finally the results related to the secondary research question.

Participants description

Initially, 52 candidates participated in the study, of whom 2 withdrew in the early stages of the study. The experiment therefore took place with 50 participants divided into 2 independent samples based on the results of the initial knowledge test. In the study, there was a choice of conducting the study in English or Italian. 17 people took part in the study in English and 33 people participated in the study in Italian (figure 2).

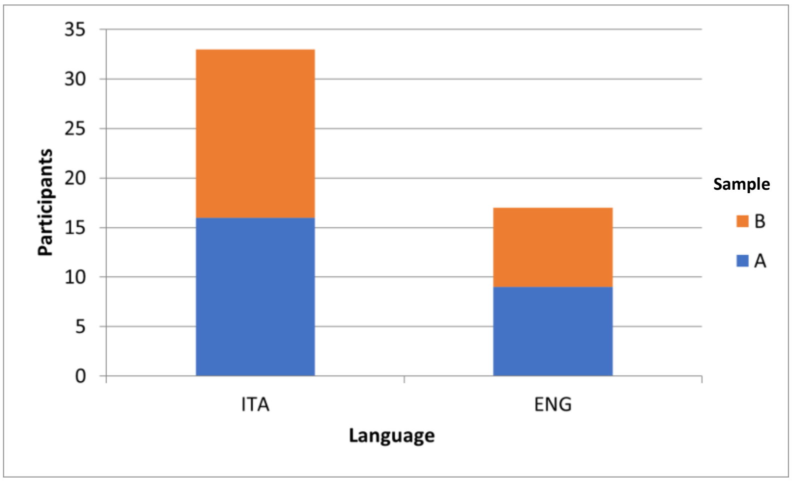

Figure 2 – Distribution of language chosen by participants

Regarding gender, in the initial sample there were 31 males and 19 females (figure 3).

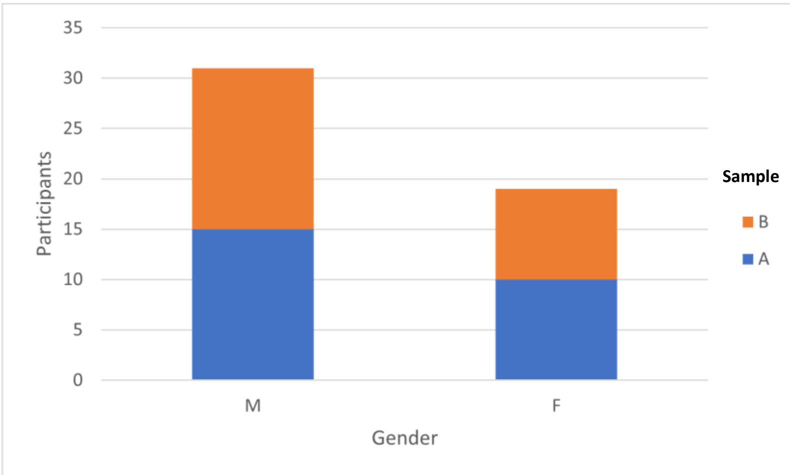

Figure 3 – Distribution of participants gender

With regard to the educational level of the participants, 38% have a high school diploma and 36% have a bachelor’s degree, the remainder 24% have a Master's Degree and 2% have a PhD (figure 4).

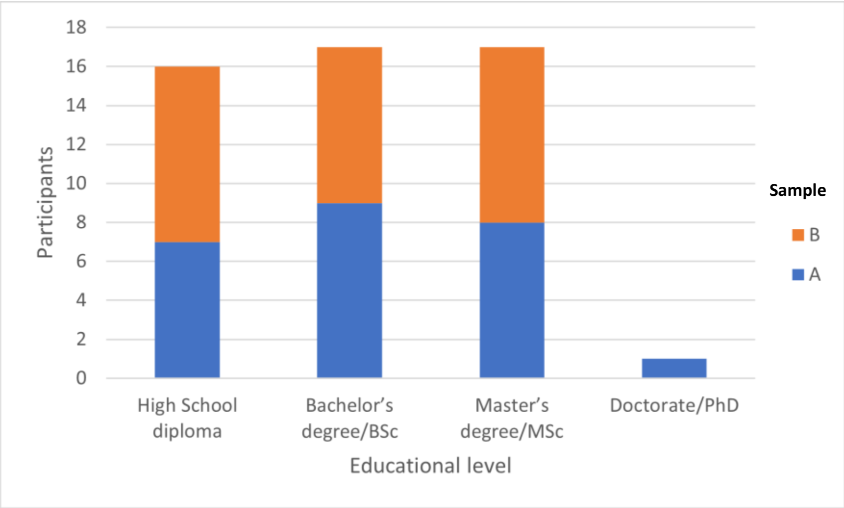

Figure 4 – Distribution of participants educational level

Concerning the employment of the participants, 21 are currently students (42%), 7 are self-employed (14%) and 22 are employed (44%) (figure 5).

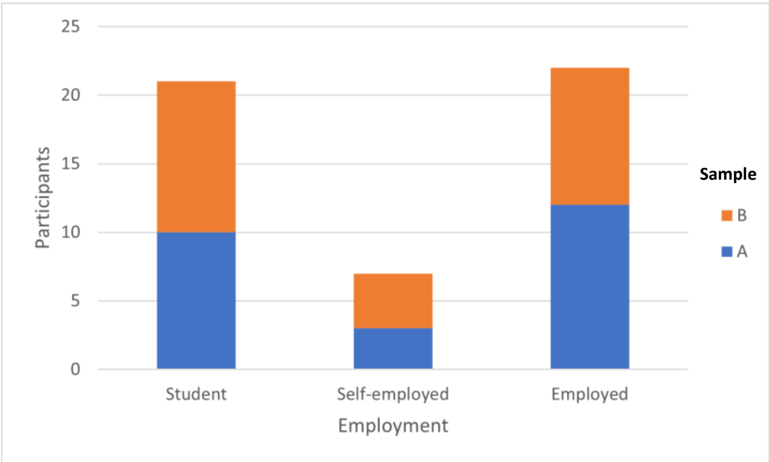

Figure 5 – Distribution of participants employment

With regard to the participants' educational background, the majority came from an economic background (44%), 26% of them had a scientific education, and then 16% came from a humanistic background. A minority group (6%) had a technological background and the remaining 8% had other types of education (figure 6).

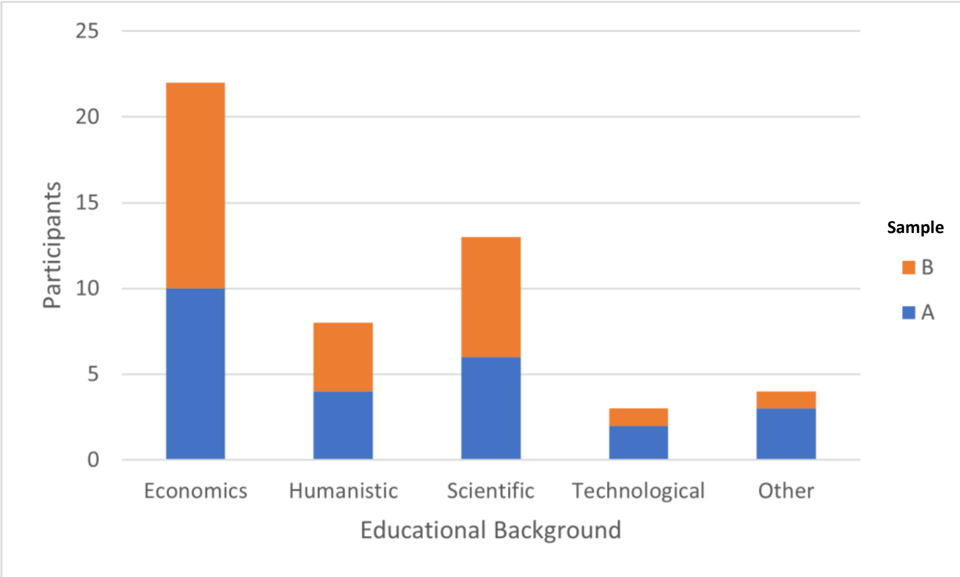

Figure 6 – Distribution of participants educational background

The mean of participants' age is 25.56 (SD = 7.68) and the median is 23 years (figure 7).

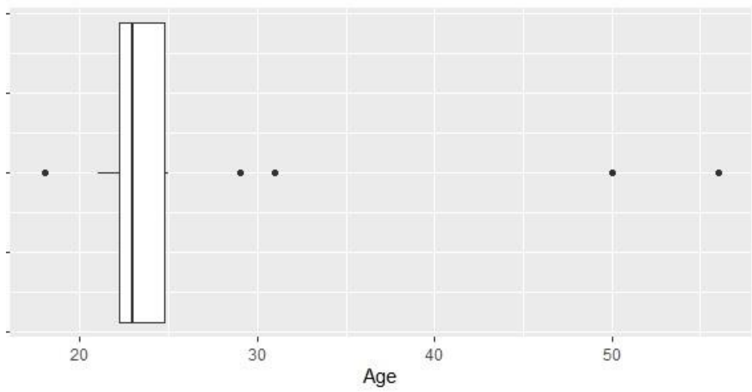

Figure 7 – Distribution of participants age

The demographic results of the study have been presented, now the results of the prior knowledges test are analysed in detail (table 5).

	Mean score (SD)	Mean time in seconds (SD)
Conceptual Knowledge	0.94 (0.65)	142s (89s)
Factual Knowledge	1.58 (0.53)	98s (57s)
Procedural Knowledge	0.94 (0.74)	98s (66s)
Final Result	3.46 (1.33)	338s (162s)

Table 5 – Summary results of prior knowledge test

The participants took an average of 5 minutes and 38 seconds to complete the knowledge test, achieving a mean score of 3.46 out of 6. In the conceptual knowledge section, the average score is 0.94 out of 2 and the average time taken to complete it was 2 minutes and 22 seconds. In factual knowledge, the average score was 1.58 out of 2 with an average test duration of 1 minute and 38 seconds. In the procedural knowledge section, the average number of points scored was 0.94 out of 2, and the section was completed in an average of 1 minute and 38 seconds.

This initial test was used to neutralise the effects of prior knowledge in the course of the experiment. As can be seen in figure 8, where the distribution of the scores obtained in the prior test knowledges is shown, sample A and sample B were created to have a similar distribution. This allows the prior knowledge in the two samples to be balanced and allows the elimination of potential bias due to the greater preparation of some components over others.

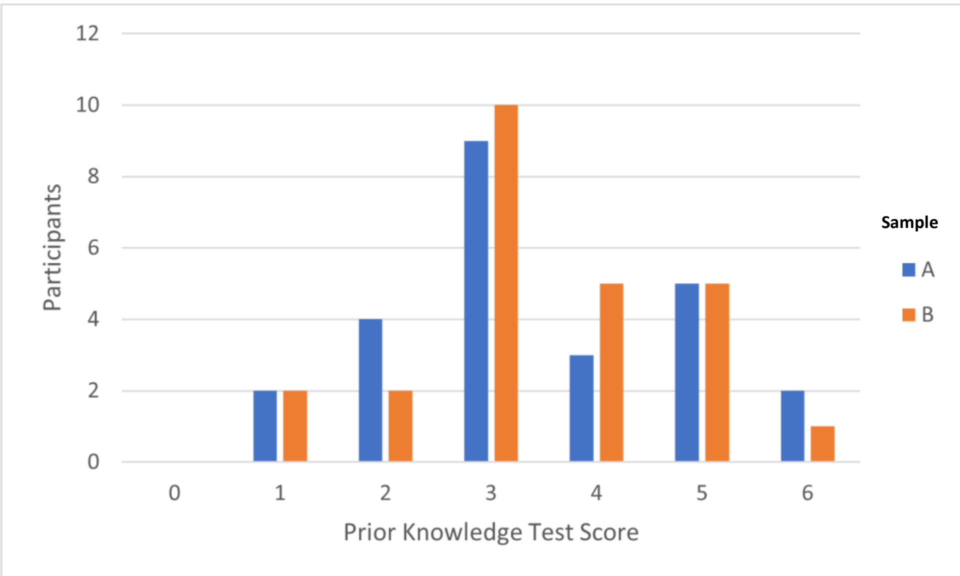

Figure 8 – Distribution of prior knowledge test score, divided by the two independent samples

Learning Style Detection

For the learning style detection, the answers given to the Honey and Mumford questionnaire were analysed. The participants on average selected 28 (SD = 9) sentences out of the 80 proposed by the questionnaire. In total, sentences belonging to the Activist category were selected 288 times, sentences belonging to the Reflector style 410 times, sentences belonging to the Theorist style 339 times and sentences belonging to the Pragmatist style 393 times.

In order to identify the learning styles of the subjects, a cluster analysis was performed using the k-means alghoritm. The best results in terms of validation of the clusters created are obtained with k=4 in terms of both internal validation and stability validation, calculated using R's package "*clValid*" (Brock et al., 2008). Measures are summarised in table 6

Validation Measures	**Results**
Internal Measures	
Connectivity	37.4806
Silhouette Width	0.0945
Dunn Index	0.2689
Stability Measures	
Average proportion of non-overlap (APN)	0.3959
Average distance (AD)	0.1782
Average distance between means (ADM)	0.0932
Figure of merit (FOM)	0.0856

Table 6 – Validation measures for cluster analysis with k = 4

Going on to interpret the resulting clusters, it can be observed that cluster 1 consists of the subjects with predominant pragmatist character (figure 9). This cluster consists of 11 observations. It can be observed that the average percentage of pragmatist responses in this cluster is 0.36 and is significantly higher than the portion of theoretical responses (Mdiff = 0.12, 95% CI [0.02, 0.22], $p < 0.01$), reflexive responses (Mdiff = 0.13, 95% CI [0.04, 0.24], $p < 0.01$) and activist responses (Mdiff = 0.17, 95% CI [0.06, 0.27], $p < 0.01$). Participants included in this cluster were therefore labelled as pragmatists.

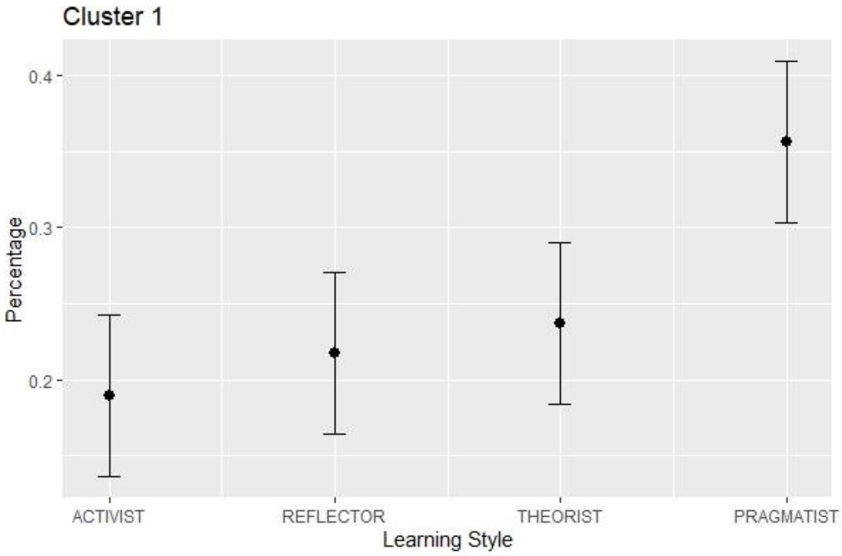

Figure 9 – Cluster 1: differences in learning styles portions

Cluster 2 consists of the subjects with predominantly reflector character (figure 10). This cluster consists of 12 observations. It can be observed that the average amount of reflector responses out of the total in this cluster is 0.41 and is significantly greater than the portion of theoretical responses (Mdiff = 0.16, 95% CI [0.09, 0.23], $p < 0.01$), pragmatist responses (Mdiff = 0.22, 95% CI [0.15, 0.29], $p < 0.01$) and activist responses (Mdiff = 0.26, 95% CI [0.20, 0.34], $p < 0.01$). Participants included in this cluster were therefore labelled as reflector.

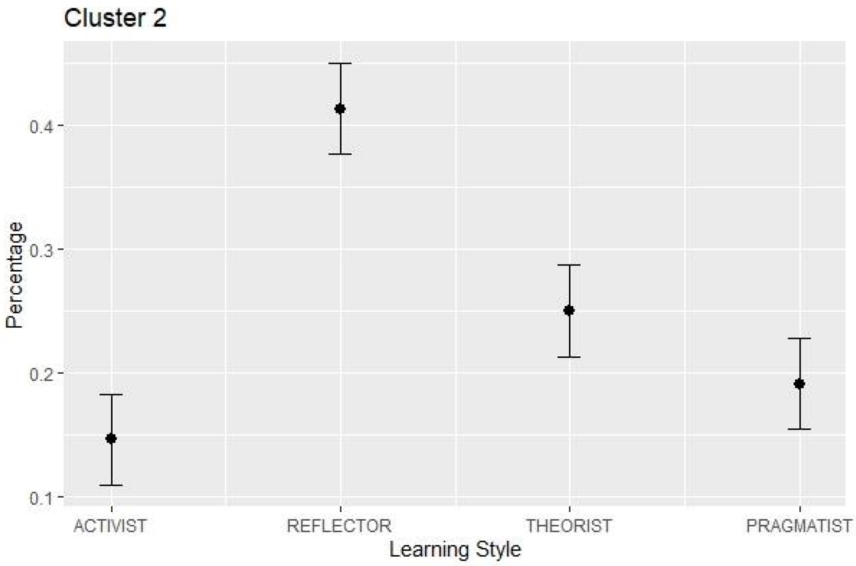

Figure 10 - Cluster 2: differences in learning styles portions

Cluster 3 includes all participants who do not have a predominant learning style in fact, as is evident from figure 11. The tests performed also show only negative significant differences, i.e., the participants included in this cluster have an activist component that is minor compared to the others. This cluster contains 16 observations, which were labelled “mixed preferences”. As also specified in the methodology, these participants at the time of sampling were placed in sample B because they had no defined preference and an error in assigning the appropriate lesson could have represented a bias.

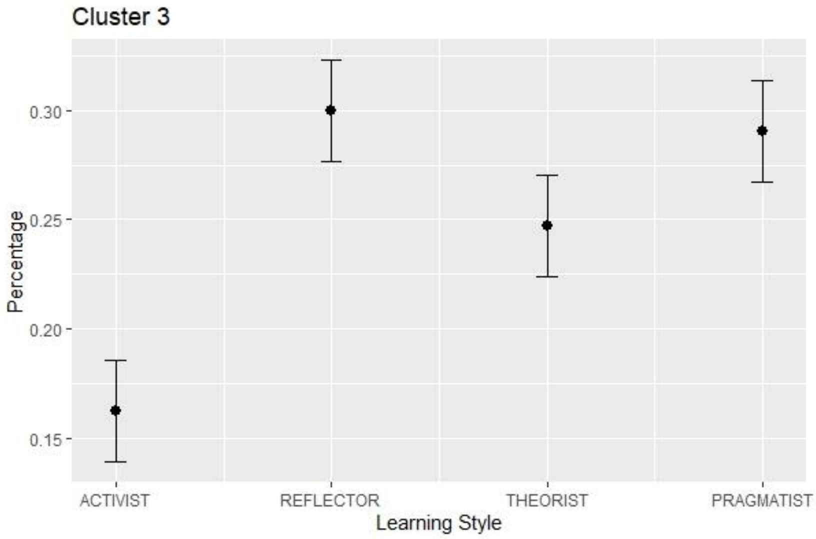

Figure 11 - Cluster 3: differences in learning styles portions

Finally, cluster 4 consists of 11 participants. The predominant learning style for this cluster is Activist as shown in figure 12. The average amount of Activist responses in relation to the total for this cluster is 0.34. This learning style is predominant because it is significantly greater than the theoretical (Mdiff = 0.14, 95% CI [0.06, 0.21], $p < 0.01$), pragmatist (Mdiff = 0.09, 95% CI [0.01, 0.17], $p < 0.01$) and reflector (Mdiff = 0.13, 95% CI [0.06, 0.20], $p < 0.01$) responses. The observations included in this cluster were therefore labelled as activists.

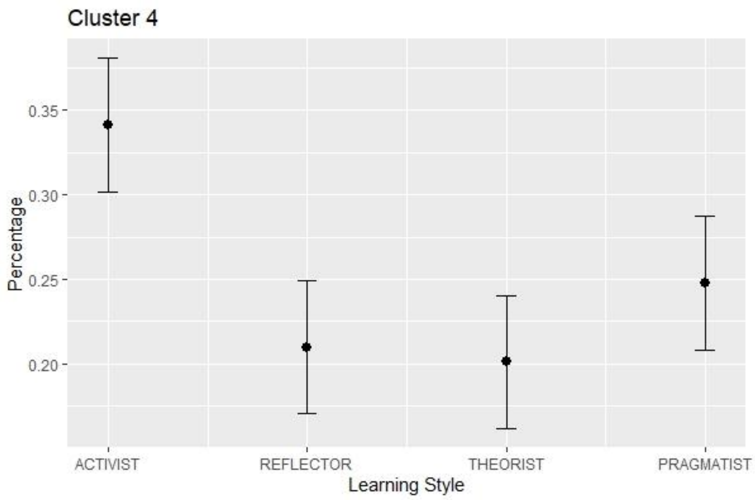

Figure 12 - Cluster 4: differences in learning styles portions

The characteristics of the participants in this study are collected in table 7. It can be observed that among the participants in this study there were no subjects with theorist as the only predominant learning style, which may represent a limitation to the power of the work because representatives of one of the four learning styles are missing.

Cluster	N. of participants	Activist %	Reflector %	Theorist %	Pragmatist %	Label
Cluster 1	11	18.9%	21.7%	23.7%	35.6%	Pragmatists
Cluster 2	12	14.6%	41.3%	25.0%	19.1%	Reflectors
Cluster 3	16	16.2%	30.0%	24.7%	29.1%	Mixed
Cluster 4	11	34.1%	21.0%	20.1%	24.8%	Activists

Table 7 – Percentage of learning styles preferences in each cluster

Performance analysis

Once the learning styles have been defined and the two samples divided, the results of the final test consisting of 8 questions are now analysed.

First of all, the overall average of the correct answers was 4.72 (SD = 1.82). The lowest score obtained was 1 and the highest score was 8. The median score is 5. In sample A, i.e. the one who was offered the lessons with an adaptive approach, the average of the correct answers was 5.16 (s.d. = 1.75). The lowest score obtained was 1 and the highest score obtained was 8. The median score is 5. In Sample B, which simulated a teaching model that does not consider learning styles, the mean of the correct answers was 4.28 (s.d. = 1.84). The lowest score obtained was 1 and the highest score obtained was 7. The median score is 4.

Population	Mean	Standard Deviation	Minimum	Maximum	Median
Sample A	5.16	1.75	1	8	5
Sample B	4.28	1.84	1	7	4
Total	4.72	1.82	1	8	5

Table 8 1 – Summary of final test results, divided by samples

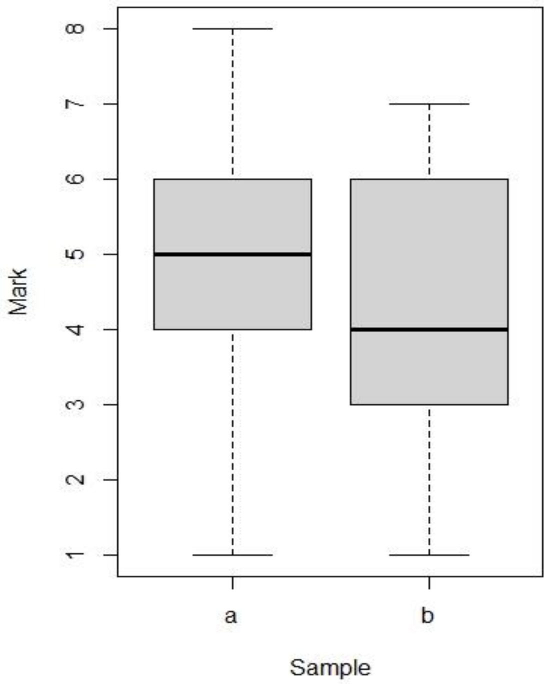

Figure 13 – Distribution of results in final tests ("Mark"), divided by sample

Before the hypothesis tests can be performed is to verify that the mean score of sample A is significantly higher than the mean score of sample B, the parametric assumptions have to be checked.

- Normality of the data: For checking the normality of the data, the Shapiro-Wilk test was adopted. The test shows that normality was not violated ($W = 0.94$, $p > 0.05$). The normality of the data can also be observed graphically in figure 14 where the variable containing the final test scores ("Mark") has a bell-shaped distribution.

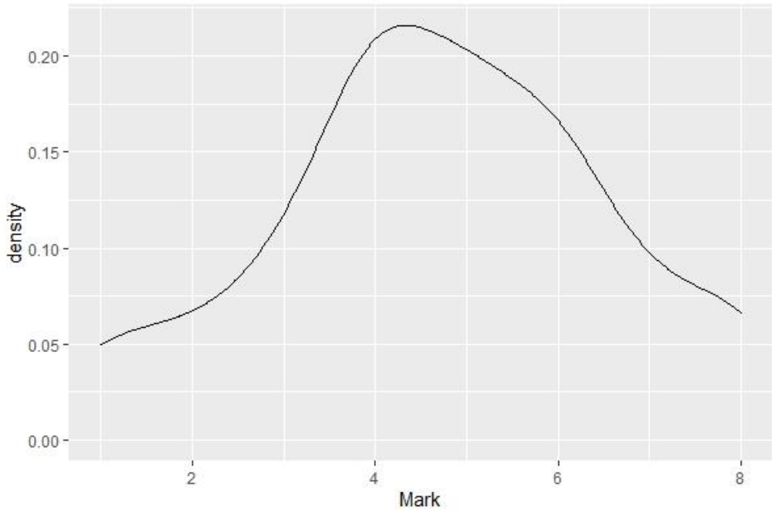

Figure 14 – Distribution of results in final test ("Mark")

- Homogeneity of variance: This assumption was verified quantitatively by means of Levene's test. The test performed was not significant [$F(1,48) = 0.39$, $p = 0.53$], which means that this assumption is also respected. This can also be recognised graphically in the boxplots proposed in figure 13.
- Equal Sample Sizes. This assumption is respected as the two samples were created considering a balance in composition in terms of size and thus both samples consist of 25 elements.

Given that all parametric assumptions are met, it is possible to proceed with the parametric tests. Specifically, having two independent samples and wanting to test the difference between the two means, it was decided to adopt the independent sample t-test.

The hypothesis tested can be represented in this way:

H_1: Adaptive e-learning average marks > Normal e-learning average marks

The result obtained from this test confirms the hypothesis formulated in the research question. In fact, the mean score of sample A, who had the adaptive e-learning experience ($M = 5.16$, $SD = 1.75$) was significantly higher than the mean score of sample B who did not have the adaptive e-learning experience ($M = 4.28$, $SD = 1.84$), [$t(48) = 1.73$, $p < 0.05$].

After obtaining a significant result in the test, its effect size was tested with Cohen's D. The result obtained is an effect size medium (d = 0.49, 95% CI [-0.09, 1.07]).

The effect size obtained lies on the borderline between a small and a medium effect size, which means that the adaptive e-learning tested in this work has a positive impact on learning performance and although the effect is not large, it can be said that the result obtained is not negligible.

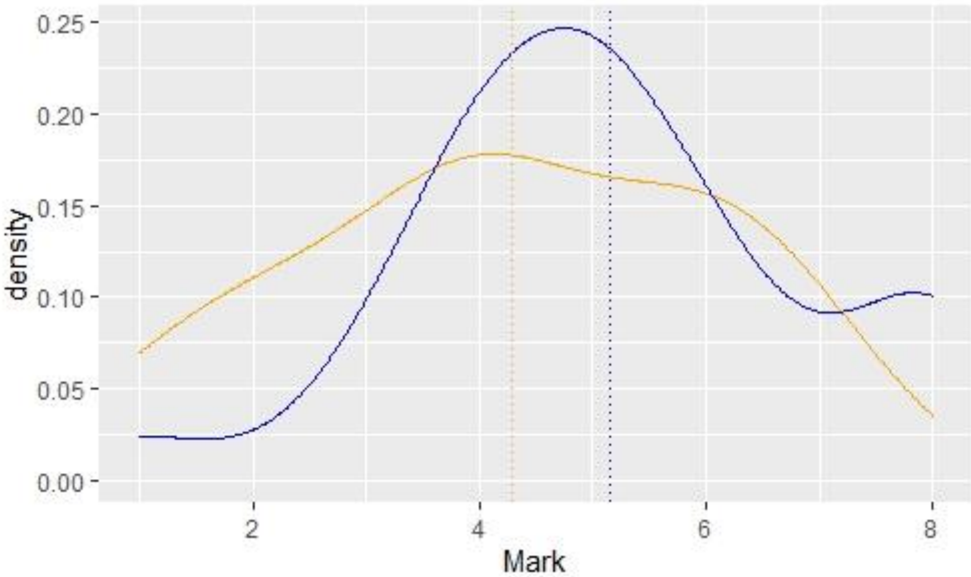

Figure 151 – Distribution of final test results ("Mark"), divided by sample. Dotted lines represents the mean of each sample. Sample A in blue and Sample B in orange.

Analysis of relationship between learning styles and interactions

To answer the second research question and look for a relationship between interactions on the prior knowledge test and learning styles, the following variables were used:

- Time spent in Conceptual Section (Time C)
- Time spent in Factual Section (Time F)
- Time spent in Procedural Section (Time P)
- Points scored in Conceptual Section (Points C)
- Points scored in Factual Section (Points F)

- Points scored in Procedural Section (Points P)
- Learning Style

The assumptions required to perform a multivariate analysis of variance (MANOVA) were checked: normality of data, homogeneity of variance, equal sample sizes and homogeneity of covariance matrices. The results are summarised in table 9. In this case, the Shapiro-Wilk tests were significant for all variables so the assumption of normality of data was violated. The Levene's tests were all non-significant, so the assumption homogeneity of variance was respected. As far as equal sample sizes are concerned, the ratio of the largest to the smallest category is less than 1.5, so it is met, and with regard to homogeneity in covariance matrices of the dependent variables, Box's M test was used, which was significant [$\chi 2$ = 65.24, df = 63, p > 0.05], meaning that this assumption is also satisfied.

	Normality of Data			**Homogeneity of Variance**		
Variable	*W*	*p-value*	*Violated?*	*F(3,48)*	*p-value*	*Violated?*
Time C	0.83	<0.05	Yes	0.74	0.53	No
Time F	0.88	<0.05	Yes	0.02	0.99	No
Time P	0.78	<0.05	Yes	0.39	0.76	No
Points C	0.79	<0.05	Yes	1.32	0.28	No
Points F	0.67	<0.05	Yes	1.78	0.16	No
Points P	0.80	<0.05	Yes	0.79	0.51	No

Table 9 – Results of parametric assumptions check

Once the assumptions have been checked, the MANOVA can be carried out despite the data not being normal, because this type of analysis is robust to allow for deviations. The dependent variables used were Time C, Time F, Time P, Points C, Points F, Points P, and as independent variable: Learning Style. The test was not significant [F(3,46) = 1.15, p = 0.30, V = 0.41], this means that the learning style variable has no significant effect on the variables analysed and therefore the learning styles do not differ in the type of interactions with performance and the duration of the prior knowledge test.

Discussion

In this chapter, the results obtained in this experiment are discussed. The first part focuses on the results obtained by studying the main research question, then in the second part the implications of the secondary research question will be assessed, and finally the problems and limitations of this work are discussed, attempting to give indications and suggestions for future research.

The main research question aimed to assess whether adaptive e-learning had a positive impact on students' learning performance, in agreement with previous studies (Essaid El Bachari & El Adnani, 2011; Latham et al. 2014; Hassan et al., 2021). The main differences with these studies are the learning style model adopted and the subject taught to which this adaptive approach was applied. The results obtained confirm those of previous studies, namely that adaptive e-learning improves learning performance even when adopting the Honey and Mumford model and teaching data analysis. The fact that the result obtained is the same regardless of the learning style model adopted and the subject taught extends the range of tests performed on adaptive e-learning and strengthens its applicability.

The effect-size obtained in the test is not particularly high and therefore reduces the power of the results obtained, at the same time, it was not so small as to render the effect obtained negligible. With regard to learning performance, this study proved to be in line with previous results and supports the effectiveness of adaptive e-learning also applied to statistical topics such as data analysis. From this perspective, the study may also have advantages for practical purposes. The experiment was carried out completely remotely and with individual study, without dynamics and contexts typical of a school environment. This makes it comparable to the individual training courses offered by training companies, which could exploit adaptive e-learning to improve their services. This scenario, however, in view of the limited number of participants and the not particularly large effect-size deserves further study.

In the secondary research question, on the other hand, the existence of relationships between learning styles and the subjects' interactions with the initial knowledge test was tested, with the aim of improving learning style detection. The results also in this case follow the existing literature, as Costa et al. (2020) found no significant effect of learning styles on interactions with the virtual learning environment. In this study, the result obtained is that of an absence of significant effects in time and performance on the initial test. Indeed, it cannot be argued that this work can give a significant contribution to learning styles detection.

Limitations and suggestion for Future Research

Talking about the limitations of this work, firstly, the number of participants should be mentioned. The number of participants achieved was below the set target and far below other similar studies in the literature (Hassan, 2021). This limitation greatly reduces the strength of the findings of this work. Another problem encountered during the course of the work concerns the treatment of participants with questionable or multiple learning styles. These could represent a bias to the study as it is not clear and defined which type of lesson suits them best. In this study, it was decided to include this category in the sample without adaptive e-learning, to try to minimise the impact of this bias and not further reduce the power of the results by decreasing the population size. The optimal solutions in this case could be twofold: to exclude this category from the study by analysing only subjects with a single predominant learning style or to study specifically the behaviour of subjects with a mixed learning style in an adaptive e-learning context.

Furthermore, participants with a theoretical learning style were absent from the sample, which may represent a limitation for the proposed analysis.

A separate discussion deserves the decision to use the comparison between two independent samples, in this specific case the decision was taken because it was the solution that allowed us to control external factors such as previous knowledges and minimise their influence. The statistical power obtained, however, is less than that which could have been achieved with two paired samples, the suggestion for future studies could be to carry out the experiment on a fixed group of participants (e.g. a class of students), adopt the paired samples technique, but carry out the experiment multiple times with different types of arguments in order to balance the uncontrollable variable of the different difficulty of the argument studied..

In general, it can be argued that studies on adaptive e-learning are yielding positive results in terms of learning performance by also using different proposed models, it can be suggested for future studies to compare the effectiveness of different learning styles models in relation to various topics and subjects.

Conclusion

E-learning has now become common practice in both university and training environments. This spread has led numerous studies to question how to increase the quality of the services offered. One area that has developed in particular is adaptive e-learning, which aims to customise teaching on the basis of the learner's characteristics.

Many studies to implement this part have exploited pedagogical theories on learning styles. These works had two main objectives, on the one hand to improve the detection of learning styles, and on the other hand to study the effectiveness of adaptive e-learning on learners.

This study tried to contribute to both areas. With regard to the effectiveness of adaptive e-learning, it was found that with the application of the Honey and Mumford model, results are positive and learning performance improves. These results confirm those obtained in the literature with the adoption of other models (e.g., Felder-Silverman model). Furthermore, the study was conducted on the teaching of a mathematical-statistical discipline such as data analysis.

It can then be argued that adapting teaching material to learners' characteristics can improve their learning performance, even in statistical subjects. The results therefore reinforce and extend, even if only partially given the statistical power obtained, the applicability of adaptive e-learning. This study may be a starting point for expanding studies on adaptive e-learning beyond the field of computer science teaching, with the adoption of different learning styles.

Unfortunately, regarding learning styles detection, no contribution can be provided, other than to confirm the absence of significant effects of learning styles in the types of interaction with the proposed material.

Finally, it can be stated that given the encouraging results in learning performance, the study can be replicated with a larger number of participants in order to increase the statistical strength of the results.

References

Al Rawashdeh, A. Z., Mohammed, E. Y., Al Arab, A. R., Alara, M., & Al-Rawashdeh, B. (2021). 'Advantages and disadvantages of using e-learning in university education: Analyzing students' perspectives'. *Electronic Journal of e-Learning*, *19*(3), 107-117.

Alonso, C. M., Gallego, D. J., & Honey, P. (1997). *Los estilos de aprendizaje: procedimientos de diagnóstico y mejora*. Bilbao, España: Mensajero.

Alsawaier, R. S. (2018) 'The effect of gamification on motivation and engagement'. *Int. J. Inform. Learn. Technol*. 35 (1). pp. 1-24

Arkorful, V., & Abaidoo, N. (2015). 'The role of e-learning, advantages and disadvantages of its adoption in higher education'. *International Journal of Instructional Technology and Distance Learning*, *12*(1), 29-42.

Barbe, W.B., Swassing, R.H., & Milone, M.N. (1979). *Teaching through modality strengths: concepts and practices*. Columbus, Ohio: Zaner-Bloser.

Bartolomé, A., Castañeda, L. & Adell, J. (2018). 'Personalisation in educational technology: the absence of underlying pedagogies'. *Int J Educ Technol High Educ* 15, 14. https://doi.org/10.1186/s41239-018-0095-0.

Behera, S.K. (2013). 'E-and M-Learning: A comparative study'. *International journal on new trends in education and their implications*, 4(3), pp.65-78.

Bennani, S., Maalel, A., and Ben Ghezala, H. (2022) 'Adaptive gamification in E-learning: A literature review and future challenges', *Comput Appl Eng Educ*. 2022; 30: 628– 642.

Bontchev, B., & Vassileva, D. (2011). 'Learning objects types dependability on styles of learning'. In *Proc. of 8th WSEAS International Conference* (pp. 227-234).

Brock, G., Pihur, V., Datta, S., & Datta, S. (2008). 'clValid: An R package for cluster validation'. *Journal of Statistical Software*, *25*, 1-22.

Cabada, R. Z., Estrada, M. L. B., & García, C. A. R. (2011). 'EDUCA: A web 2.0 authoring tool for developing adaptive and intelligent tutoring systems using a Kohonen network'. *Expert Systems with Applications*, *38*(8), 9522-9529.

Caballé, S. and Clarisó, R., (2016). 'Formative assessment, learning data analytics and gamification: In ICT education'. Morgan Kaufmann.

Clark, J.D. (2007). *Learning and Teaching in the Mobile Learning Environment of the Twenty-First Century*. Austin, Texas: Austin Community College.

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Routledge.

Costa, R. D., Souza, G. F., Valentim, R. A., & Castro, T. B. (2020). 'The theory of learning styles applied to distance learning'. *Cognitive Systems Research*, *64*, 134-145.

El Aissaoui, O., El Madani, Y. E. A., Oughdir, L., & El Allioui, Y. (2019). 'Combining supervised and unsupervised machine learning algorithms to predict the learners' learning styles'. *Procedia computer science*, *148*, 87-96.

El Bachari, E., & El Hassan Abelwahed, M. E. A. (2011). 'E-learning personalization based on dynamic learners'preference'.

Elnaga, A., & Imran, A. (2013). 'The effect of training on employee performance'. *European journal of Business and Management*, *5*(4), 137-147.

Eurostat, (2022). 'Educational attainment statistics' . Available at: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Educational_attainment_statistics#Development_of_educational_attainment_levels_over_time.

Felder, R. M., & Silverman, L. K. (1988). Learning and teaching styles in engineering education. *Engineering education*, *78*(7), 674-681.

Gašević, D., Dawson, S., Rogers, T., & Gasevic, D. (2016). 'Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success'. *The Internet and Higher Education*, *28*, 68-84.

GDPR. (2016). REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679

Germanakos, P., Tsianos, N., Lekkas, Z., Mourlas, C., & Samaras, G. (2008). ‘Capturing essential intrinsic user behaviour values for the design of comprehensive web-based personalized environments’. *Computers in Human Behavior*, *24*(4), 1434-1451.

Hassan, M. A., Habiba, U., Majeed, F. & Shoaib, M. (2021) ‘Adaptive gamification in e-learning based on students’ learning styles’, *Interactive Learning Environments*, 29:4, 545-565, DOI: 10.1080/10494820.2019.1588745

Hassenburg, A. (2009). ‘Distance education versus the traditional classroom’. *Berkeley Scientific Journal*, *13*(1).

Honey, P, & Mumford, A. (1982). *The manual of learning styles*. Maidenhead: Peter Honey.

Honey, P., & Mumford, A. (1989). *Learning styles questionnaire*. Organization Design and Development, Incorporated.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112, p. 18). New York: springer.

Jones, E. (2017). ‘One size fits all? Multiple intelligences and legal education’. *The Law Teacher*, *51*(1), 56-68.

Keller, C., & Cernerud, L. (2002). ‘Students’ perceptions of e-learning in university education’. *Journal of Educational Media*, *27*(1-2), 55-67.

Kolb, D. A. (1984). *Experiential learning: Experience as the source of learning and development*. FT press.

Krathwohl, D. R. (2002). ‘A Revision of Bloom's Taxonomy: An Overview’, *Theory Into Practice*, 41:4, 212-218, DOI: 10.1207/s15430421tip4104_2

Kulaglić, S., Mujačić, S., Serdarević, I. K., & Kasapović, S. (2013). ‘Influence of learning styles on improving efficiency of adaptive educational hypermedia systems’. In *2013 12th International Conference on Information Technology Based Higher Education and Training (ITHET)* (pp. 1-7). IEEE.

Kumar Basak, S., Wotto, M., & Belanger, P. (2018). ‘E-learning, M-learning and D-learning: Conceptual definition and comparative analysis’. *E-learning and Digital Media*, *15*(4), 191-216.

Kurilovas, E., Kubilinskiene, S., & Dagiene, V. (2014). 'Web 3.0–Based personalisation of learning objects in virtual learning environments'. *Computers in Human Behavior*, *30*, 654-662.

Latham, A., Crockett, K., & McLean, D. (2014). An adaptation algorithm for an intelligent natural language tutoring system. *Computers & Education*, *71*, 97-110.

Levene, H. (1960). Robust tests for equality of variances. In *Contributions to Probability and Statistics*, I. Olkin, S. G. Ghurye, W. Hoeffding, W. G. Madow, and H. B. Mann (eds), 278 – 292. Stanford, California : Stanford University Press

Lin, C. F., Yeh, Y. C., Hung, Y. H., & Chang, R. I. (2013). 'Data mining for providing a personalized learning path in creativity: An application of decision trees'. *Computers & Education*, *68*, 199-210.

Lovaglio, P.G., Cesarini, M., Mercorio, F., Mezzanzanica, M. (2018). 'Skills in demand for ICT and statistical occupations: Evidence from web-based job vacancies.' *Stat Anal Data Min: The ASA Data Sci Journal.* 11:78–91

Malik, S., & Rana, A. (2020). 'E-Learning: Role, Advantages, and Disadvantages of its implementation in Higher Education'. *JIMS8I-International Journal of Information Communication and Computing Technology*, *8*(1), 403-408.

Pagram, P. and Pagram, J. (2006). 'Issues In E-Learning: A Thai Case Study'. *The Electronic Journal of Information Systems in Developing Countries*, 26: 1-8. https://doi.org/10.1002/j.1681-4835.2006.tb00175.x.

Paireekreng, W., & Prexawanprasut, T. (2015). 'An integrated model for learning style classification in university students using data mining techniques'. In *2015 12th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)* (pp. 1-5). IEEE.

Pashler, H., McDaniel, M., Rohrer, D., & Bjork, R. (2008). 'Learning styles: Concepts and evidence'. *Psychological science in the public interest,* 9(3), 105-119.

Prodanov, C. C., & De Freitas, E. C. (2013). *Metodologia do trabalho científico: métodos e técnicas da pesquisa e do trabalho acadêmico-2ª Edição*. Editora Feevale.

Radović-Marković, M. (2010). ‘Advantages and disadvantages of e-learning in comparison to traditional forms of learning’. *Annals of the University of Petroşani, Economics*, *10*(2), 289-298.

Raykova, M., Kostadinova, H., & Totkov, G. (2011). ‘Adaptive test system based on revised Bloom's taxonomy’. In *Proceedings of the 12th International Conference on Computer Systems and Technologies* (pp. 504-509).

Razali, N. M., & Wah, Y. B. (2011). Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests. *Journal of statistical modeling and analytics*, *2*(1), 21-33.

Riener, C., & Willingham, D. (2010). ‘The myth of learning styles’. *Change: The magazine of higher learning*, *42*(5), 32-35.

Rohrer, D., & Pashler, H. (2012). Learning Styles: ‘Where's the Evidence?’. *Online Submission*, *46*(7), 634-635.

Sangrà, A., Vlachopoulos, D., & Cabrera, N. (2012). ‘Building an inclusive definition of e-learning: An approach to the conceptual framework’. *International Review of Research in Open and Distributed Learning*, *13*(2), 145-159.

Saraswathy, R. (2019) ‘Relationship between Learning Strategies in Mathematics and Learning Styles among IX Standard Students’, *i-Manager's Journal on School Educational Technology*, vol. 15, no. 1, pp. 36-44.

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, *52*(3/4), 591-611.

Shute, V., & Towle, B. (2018). ‘Adaptive e-learning’. In *Educational psychologist* (pp. 105-114). Routledge.

Siddiqi, M. H., Alsayat, A., Alhwaiti, Y., Azad, M., Alruwaili, M., Alanazi, S., ... & Khan, A. (2022). ‘A Precise Medical Imaging Approach for Brain MRI Image Classification’. *Computational Intelligence and Neuroscience*.

Tayebinik, M., & Puteh, M. (2012). ‘Blended Learning or E-learning?’. *International Magazine on Advances in Computer Science and Telecommunications*, 3(1), 103-110.

Thalmann, S. (2014). ‘Adaptation criteria for the personalised delivery of learning materials: A multi-stage empirical investigation’. *Australasian Journal of Educational Technology*, *30*(1).

Truong, H. M. (2016). ‘Integrating learning styles and adaptive e-learning system: Current developments, problems and opportunities’. *Computers in human behavior*, *55*, 1185-1193.

Vita, G. D. (2001). ‘Learning styles, culture and inclusive instruction in the multicultural classroom: A business and management perspective’. *Innovations in Education and Teaching International*, *38*(2), 165-174

Yang, S., Tian, H., Sun, L., & Yu, X. (2019). ‘From one-size-fits-all teaching to adaptive learning: the crisis and solution of education in the era of AI’. In *Journal of Physics: Conference Series* (Vol. 1237, No. 4, p. 042039). IOP Publishing.

Yang, T. C., Hwang, G. J., & Yang, S. J. H. (2013). ‘Development of an adaptive learning system with multiple perspectives based on students' learning styles and cognitive styles’. *Journal of Educational Technology & Society*, *16*(4), 185-200.

Yousef, D.A., (2016). ‘Learning styles preferences of statistics students: A study in the Faculty of Business and Economics at the UAE University’. *Quality Assurance in Education*.

Wang, M., Ran, W., Liao, J. and Yang, S.J. (2010). ‘A performance-oriented approach to e-learning in the workplace’. *Journal of Educational Technology & Society*, *13*(4), pp.167-179.

Appendices

Appendix I – Honey and Mumford questionnaire and demographic questions

	Learning style	Sentence	Ticked
1	*Theorist*	I have strong beliefs about what is right and wrong, good and bad.	34
2	*Activist*	I often act without considering the possible consequences.	5
3	*Theorist*	I tend to solve problems using a step-by-step approach, avoiding any 'flights of fancy'.	28
4	*Activist*	I believe that formal procedures and policies restrict people.	20
5	*Pragmatist*	I have a reputation for saying what I think, simply and directly.	30
6	*Activist*	I often find that actions based on feelings are as sound as those based on careful thought and analysis.	8
7	*Reflector*	I like the sort of work where I have time for thorough presentation and implementation.	16
8	*Theorist*	I regularly question people about their basic assumptions.	12
9	*Pragmatist*	What matters most is whether something works in practice.	29
10	*Activist*	I actively seek out new experiences.	29
11	*Pragmatist*	When I hear about a new idea or approach I immediately start working out how to apply it in practice.	22
12	*Theorist*	I am keen on self-discipline such as watching my diet, taking regular exercise, sticking to routine.	19
13	*Reflector*	I take pride in doing a thorough job.	20
14	*Theorist*	I get on best with logical people and less well with spontaneous 'irrational' people.	10
15	*Reflector*	I take care over the interpretation of data available to me and avoid jumping to conclusions.	8
16	*Reflector*	I like to reach a decision carefully after weighing up many alternatives.	21
17	*Activist*	I'm attracted more to novel, unusual ideas than practical ones.	9
18	*Theorist*	I don't like disorganised things and prefer to fit things into a coherent pattern.	26
19	*Pragmatist*	I accept and stick to laid down procedures and policies as long as I regard them as an efficient way of getting the job done.	10

20	*Theorist*	I like to relate my actions to general principles.	18
21	*Pragmatist*	In discussions I like to get straight to the point.	30
22	*Theorist*	I tend to have distant rather formal relationships with the people at work.	6
23	*Activist*	I thrive on the challenge of tackling something new and different.	32
24	*Activist*	I enjoy fun-loving spontaneous people.	40
25	*Reflector*	I pay meticulous attention to detail before coming to a conclusion.	21
26	*Theorist*	I find it difficult to produce ideas on impulse.	6
27	*Pragmatist*	I believe in coming to the point immediately.	12
28	*Reflector*	I am careful not to jump to conclusions too quickly.	25
29	*Reflector*	I prefer to have as many sources of information as possible – the more data to think over the better.	32
30	*Theorist*	Flippant people who don't take things seriously usually irritate me.	30
31	*Reflector*	I listen to other people's point of view before putting my own forward.	24
32	*Activist*	I tend to be open about how I'm feeling.	13
33	*Reflector*	In discussions I enjoy watching the manoeuvring of the other participants.	30
34	*Activist*	I prefer to respond to events on a spontaneous, flexible basis rather than plan things in advance.	12
35	*Pragmatist*	I tend to be attracted to techniques such as network analysis, flow charts, branching programmes, contingency planning.	9
36	*Reflector*	It worries me if I have to rush a piece of work to meet a tight deadline.	12
37	*Pragmatist*	I tend to judge people's ideas on their practical merits.	14
38	*Activist*	Quiet, thoughtful people tend to make me feel uneasy.	3
39	*Reflector*	I often get irritated by people who want to rush things.	26
40	*Activist*	It is more important to enjoy the present moment than to think about the past or future.	21
41	*Reflector*	I think that decisions based on thorough analysis of all the information are sounder than those based on intuition.	17
42	*Theorist*	I tend to be a perfectionist.	19
43	*Activist*	In discussions I usually pitch in with lots of spontaneous ideas.	9
44	*Pragmatist*	In meetings I put forward practical, realistic ideas.	21

45	*Activist*	More often that not, rules are there to be broken.	11
46	*Reflector*	I prefer to stand back from a situation and consider all the perspectives.	17
47	*Theorist*	I can often see inconsistences and weaknesses in other people's arguments.	29
48	*Activist*	On balance I talk more than I listen.	12
49	*Pragmatist*	I can often see better more practical ways to get things done.	23
50	*Pragmatist*	I think written reports should be short, punchy and to the point.	30
51	*Theorist*	I believe that rational, logical thinking should win the day.	18
52	*Reflector*	I tend to discuss specific things with people rather than engaging in social discussion.	11
53	*Pragmatist*	I like people who approach things realistically rather than theoretically.	32
54	*Pragmatist*	In discussions I get impatient with with irrelevancies and digressions.	17
55	*Reflector*	If I have a report to write I tend to produce lots of drafts before settling on the final version.	9
56	*Pragmatist*	I am keen to try things out to see if they work in practice.	22
57	*Theorist*	I am keen to reach answers via a logical approach.	19
58	*Activist*	I enjoy being the one who talks a lot.	6
59	*Pragmatist*	In discussions, I often find I am the realist, keeping people to the point and avoiding wild speculations.	21
60	*Reflector*	I like to ponder many alternatives before making up my mind.	24
61	*Theorist*	In discussions with people I often find I am the most dispassionate and objective.	13
62	*Reflector*	In discussions I'm more likely to adapt a 'low profile' than to the take the lead and do most of the talking.	15
63	*Theorist*	I like to be able to relate current actions to a longer term.	14
64	*Activist*	When things go wrong I am happy to shrug it off and 'put it down to experience'.	9
65	*Pragmatist*	I tend to reject wild, spontaneous ideas as being impractical.	6
66	*Reflector*	It's best to think carefully before taking action.	27
67	*Reflector*	On balance I do the listening rather than the talking.	23

68	*Theorist*	I tend to be tough on people who find it difficult to adopt a logical approach.	9
69	*Pragmatist*	Most times I believe the end justifies the means.	18
70	*Pragmatist*	I don't mind hurting people's feelings so long as the job gets done.	10
71	*Activist*	I find the formality of having specific objectives and plans stifling.	6
72	*Activist*	I'm usually one of the people who puts life into a party.	17
73	*Pragmatist*	I do whatever is expedient to get the job done.	26
74	*Activist*	I quickly get bored with methodical, detailed work.	15
75	*Theorist*	I am keen on exploring the basic assumptions, principles and theories underpinning things and events.	11
76	*Reflector*	I'm always interested to find out what people think.	32
77	*Theorist*	I like meetings to be run on methodical lines, sticking to laid-down agenda etc.	12
78	*Theorist*	I steer clear of subjective or ambiguous topics.	6
79	*Activist*	I enjoy the drama and excitement of a crisis situation.	11
80	*Pragmatist*	People often find me insensitive to their feelings.	11

Table 20 - 80-items Honey and Mumford Questionnaire (Honey & Mumford, 1989)

Background and Demographic questions

1. **Indicate your age (in number)**
2. **Indicate your education level (choose higher course completed)**
    a. Doctorate/PHD
    b. Master’s degree/MSc
    c. Bachelor’s degree/MSc
    d. High School diploma
    e. Middle School diploma

3. **Indicate your current employment:**
    a. Student
    b. Unemployed
    c. Employed
    d. Self-Employed
    e. Retired
4. **Indicate your educational background:**
    a. Scientific
    b. Humanistic
    c. Economics
    d. Technological
    e. Other
5. **Indicate your gender:**
    a. Male
    b. Female
    c. Prefer not to say

Appendix II – Prior Knowledge test

This test is divided in three sections, with two multiple choice questions per each. In bold characters, the correct answers are highlighted.

Factual Knowledges

6. **In the context of data analysis, which of the following phrases best describes the concept of predictive analysis:**
    a. It is a process in which, through the use of statistical tools, historical data is analysed in order to describe what happened in the past.
    b. **It is a process in which, through the use of statistical tools and machine learning models, historical data is analysed in order to make predictions about what may happen in the future.**
    c. It is the process of analysing the possible scenarios that may occur after an investment in the next five years.

d. It is the set of activities carried out in order to predict climate change.

7. **In a model, if the dependent variable (the variable we want to predict), is categorical, we are facing a problem of:**
    a. **Classification**
    b. Regression
    c. Clustering
    d. Dimension Reduction

Conceptual Knowledges

8. **If in a linear regression model, the dependent variable is 'house price' and the coefficient of the independent variable 'area' is 100, how can this be interpreted?**
    a. The price changes every 100 units of area.
    b. As the price increases by 1 unit, the area increases by 100 units.
    c. Forecasts can only be made if the area is greater than 100 units.
    d. **As the area increases by 1 unit, the price increases by 100 units**.

9. **If in a problem you are asked to model whether a customer will be able to repay a mortgage or not (binary variable), which is NOT a suitable tool to answer the question?**
    a. Logistic Regression
    b. Decision tree classifier
    c. **Linear Regression**
    d. Neural Network

Procedural Knowledges

10. **Which of the following is a correct sequence for performing a predictive analysis using machine learning tools:**
    a. **Collect data, prepare and clean data, train model, evaluate model, make prediction.**
    b. Make prediction, prepare and clean data, train model, evaluate model, collect data.

c. Train model, collect data, prepare and clean data, make prediction, evaluate model.
d. Prepare and clean data, collect data, train model, evaluate model, make prediction.

11. **At what stage in the analysis of a linear regression should diagnostic plots (i.e. representation of residual characteristics) be consulted?**
    a. At the data cleaning and preparation stage
    b. Before training the model
    c. After the model has been trained on the data
    d. Those should not be consulted

Appendix III – Final Test

This test is composed by 8 multiple choice questions. The correct answers are highlighted in bold characters.

1. **What are the main shrinkage methods?**
    a. Ridge Regression and Lasso Regression
    b. Linear Regression and Logistic Regression
    c. Forward and backward stepwise selection
    d. Ridge Regression and Stepwise selection

2. **Which of the following is not a criterion that can be used in subset selection methods?**
    a. Akaike's Information Criteria (AIC)
    b. Bayesian Information Criteria (BIC)
    c. Adjusted R^2
    d. Lasso

3. **Why is there no dependent variable (target) in unsupervised learning methods?**
    a. Because they serve to make predictions
    b. Because they should be used after making predictions about the target variable

c. **Because they are exploited to discover hidden patterns and subcategories among the variables**
d. In unsupervised learning methods there is a dependent variable (target)

4. **Why is subset selection useful when you have a dataset with many variables?**
   a. To save computer memory
   b. **Because removing insignificant variables improves the interpretability of the model**
   c. Because several models can be run simultaneously
   d. Because you want to keep all variables but cancel out the incidence of non-significant ones

5. **The following table represents a stage of a backward selection. Considering that the selection criterion used in this case is AIC, how would you analyse the result?**

   (N.B. the dependent variable in the model is 'medv' and all other variables are independent. In the table below. In the table, the row 'None' represents the complete model, while the other rows represent the complete model with the exclusion of the indicated variable)

```
medv ~ crim + zn + indus + chas + nox + rm + dis + rad + tax +
    ptratio + black + lstat

          Df Sum of Sq   RSS    AIC
- indus    1      2.52 11081 1636.5
<none>                 11079 1642.6
- chas     1    219.91 11299 1646.3
- tax      1    242.24 11321 1647.3
- crim     1    243.20 11322 1647.3
- zn       1    260.32 11339 1648.1
- black    1    272.26 11351 1648.7
- rad      1    481.09 11560 1657.9
- nox      1    520.87 11600 1659.6
- ptratio  1   1200.23 12279 1688.4
- dis      1   1352.26 12431 1694.6
- rm       1   1959.55 13038 1718.8
- lstat    1   2718.88 13798 1747.4
```

   a. **Subset selection process is incomplete because removing a variable we can achieve a lower AIC**

b. Subset selection process is incomplete because removing a variable we can achieve a model with a higher AIC
c. Subset selection process is completed, and the final model contains only the variable 'indus'.
d. Subset selection process is completed, and the final model contains no variables

6. You are asked to perform an analysis of data from a clothing company in order to:

   1 - Find out whether consumer clusters exist among their customers and the composition of these

   2 - Predict whether colour influences sales performance and if so how

   What kind of analysis would you conduct?

   a. For the task 1 unsupervised learning and for the task 2 supervised learning
   b. For the task 1 supervised learning and for the task 2 unsupervised learning
   c. For both supervised learning
   d. For both unsupervised learning

7. Consider a dataset with the following variables:

   "medv" = median value of a house in dollars

   "crim" = index of crimes committed in the area per inhabitant

   "age" = number of housing units in the area built before 1940

   And the following model:

$$medv = 29800.67 - 311.8 * crim - 89.5 * age$$

   Identify which of the following interpretations is NOT correct.

   a. When the crime index increases by one point the value of houses decreases by \$311.8
   b. The more pre-1940 housing units there are in the area, the more the house value decreases
   c. The more pre-1940 housing units there are in the area, the more the house value increases

d. If the crime rate is 0 and there are no pre-1940 houses in the area, the median value of a house is $29800.67

8. In the ridge regression, if the penalisation parameter is a number, can a coefficient be equal to 0?

a. Yes
b. No coefficients are considered in this regression
c. No, they are between 1 and 100
d. No, they can only go close to zero

Appendix IV – Lectures preview

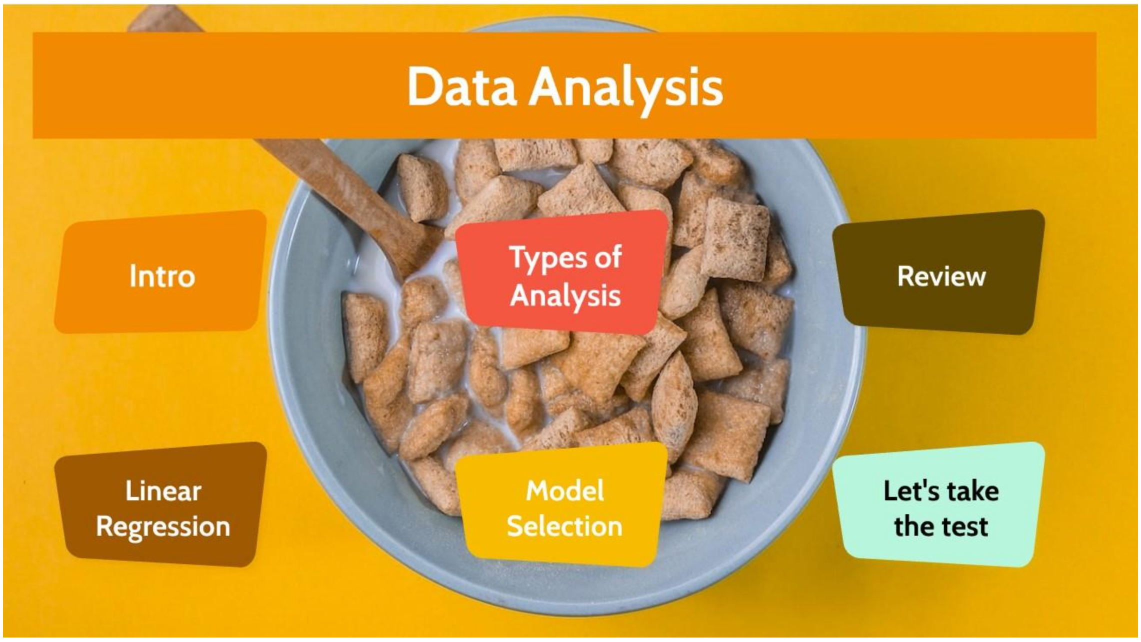

Figure 162 – Activist Lesson preview

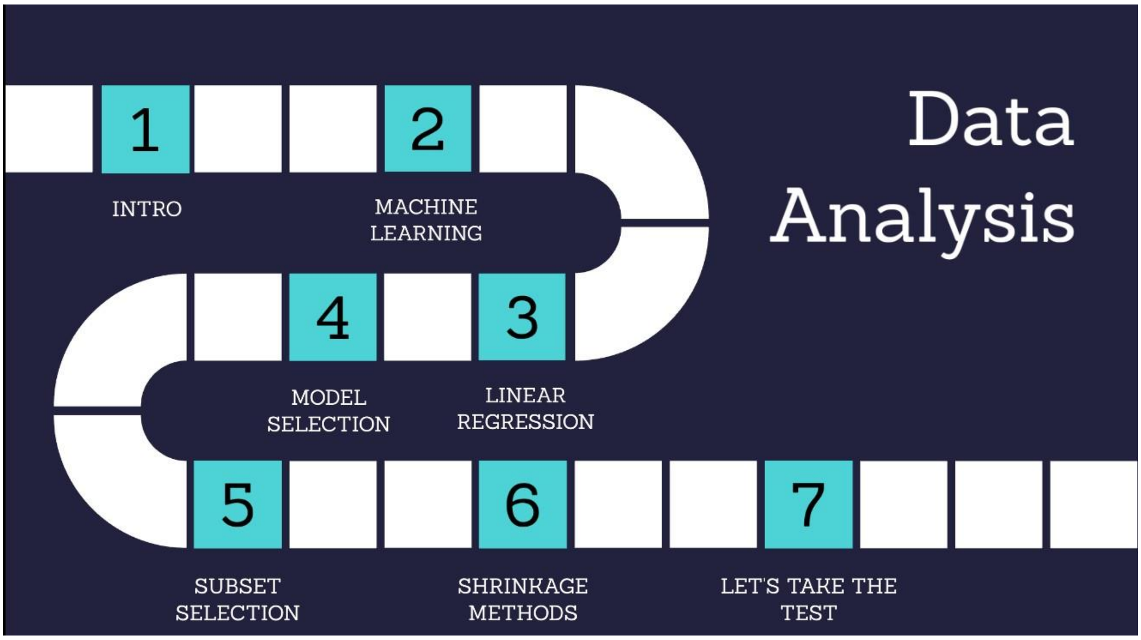

Figure 173 – Pragmatist lesson preview

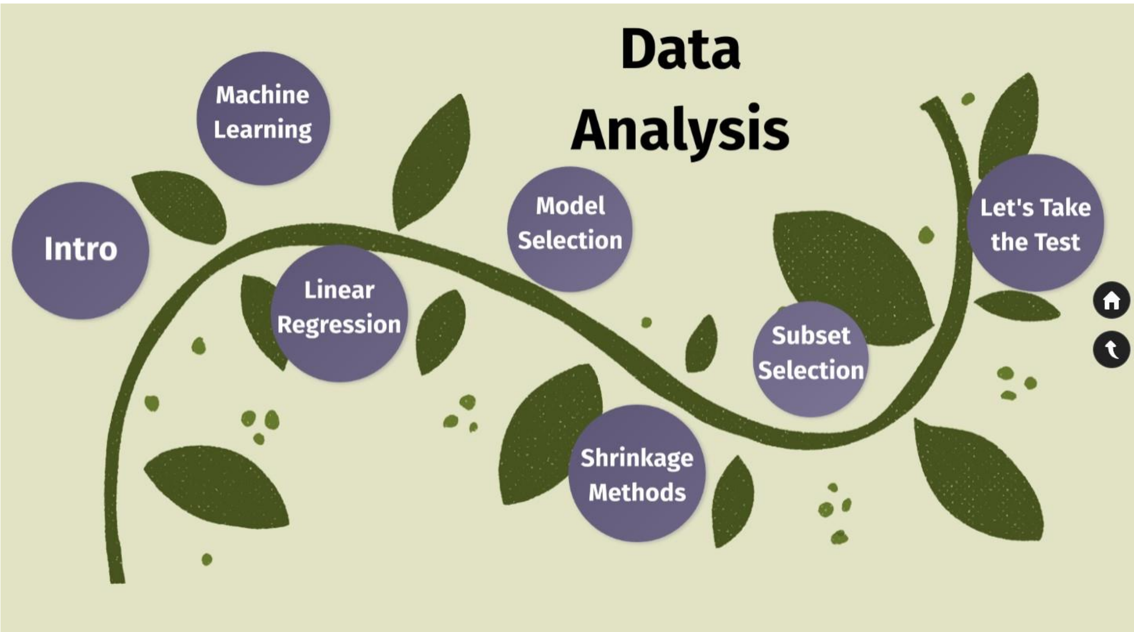

Figure 18 – Theorist lesson preview

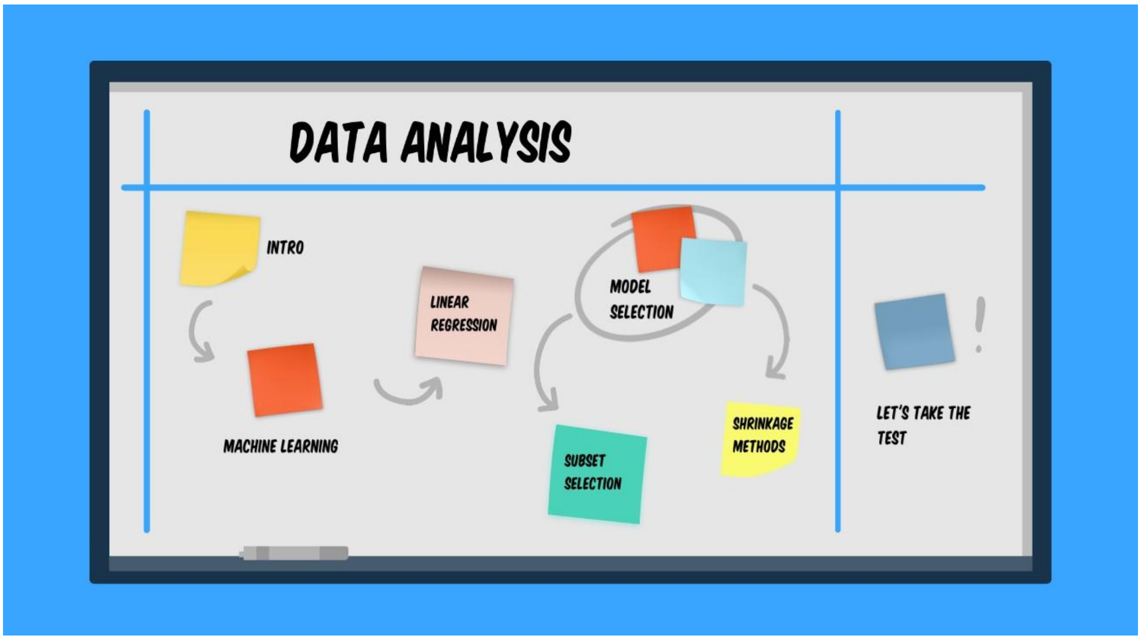

Figure 19 – Reflector Lesson preview