# AN INTERVENTION FRAMEWORK FOR DESIGN AND DEVELOPMENT OF INTERACTIVE VIDEO LECTURES BASED ON VIDEO VIEWING BEHAVIORS: A LEARNING ANALYTICS APPROACH

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF INFORMATICS OF
MIDDLE EAST TECHNICAL UNIVERSITY
BY

OZAN RAŞİT YÜRÜM

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
IN
THE DEPARTMENT OF INFORMATION SYSTEMS

SEPTEMBER 2021

Approval of the thesis:

**AN INTERVENTION FRAMEWORK FOR DESIGN AND DEVELOPMENT OF INTERACTIVE VIDEO LECTURES BASED ON VIDEO VIEWING BEHAVIORS: A LEARNING ANALYTICS APPROACH**

Submitted by **OZAN RAŞİT YÜRÜM** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy in Information Systems Department, Middle East Technical University** by,

**Date:** **07.09.2021**

**I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.**

**Name, Last name : Ozan Raşit Yürüm**

**Signature :** _____

## ABSTRACT

### AN INTERVENTION FRAMEWORK FOR DESIGN AND DEVELOPMENT OF INTERACTIVE VIDEO LECTURES BASED ON VIDEO VIEWING BEHAVIORS: A LEARNING ANALYTICS APPROACH

YÜRÜM, OZAN RAŞİT

Ph.D., Department of Information Systems

Supervisor: Prof. Dr. İbrahim Soner Yıldırım

Co-Supervisor: Assoc. Prof. Dr. Tuğba Taşkaya Temizel

September 2021, 188 Pages

Video-based learning has become very popular together with the increase in the number of massive open online courses (MOOCs). Learners interacting with video naturally exhibit viewing behaviors based on clickstream interactions such as pause, forward, and backward. These are important inputs to predict learning variables and to develop an intervention framework. However, there is no guideline or method on how to use these behavioral patterns in favor of learners. The purpose of this study is to investigate learners' video clickstream behaviors, and to use them to develop a framework in order to increase the impact of video lectures. Accordingly, three experiments were performed. In the first one, a pilot study with 22 students was conducted. In the second one, a more comprehensive experiment was performed with 29 students to develop the framework. In the third one, an experiment based on one group pretest-posttest design was conducted with 18 students to evaluate the framework. These studies reveal important insights in terms of both the framework and comprehensive learning analytics. The results for the framework show that students' satisfaction increases significantly in interactive videos created based on the framework when motivation is controlled. In addition, students' frequency to go back to important points decreases significantly and students' frequency to skip unimportant points increases significantly in interactive videos. Apart from the framework, the results show that the mean of backward speeds, number of pauses, and number/percentage of backwards are the most important indicators in predicting students' grades. Furthermore, the study also shows that video clickstream behaviors are able to predict grade whereas the reasons behind the behaviors are also necessary to have to be able to predict satisfaction successfully. However, video clickstream behaviors and their reasons are not enough to predict engagement significantly.



**Keywords:** Intervention Framework, Learning Analytics, Educational Data Mining, Video-Based Learning, Interactive Videos

# ÖZ

## VİDEO İZLEME DAVRANIŞLARINA DAYALI ETKİLEŞİMLİ VİDEO DERSLERİN TASARIMI VE GELİŞTİRİLMESİ İÇİN BİR MÜDAHALE ÇERÇEVESİ: BİR ÖĞRENME ANALİTİĞİ YAKLAŞIMI

YÜRÜM, OZAN RAŞİT

Doktora, Department of Information Systems

Tez Danışmanı: Prof. Dr. İbrahim Soner Yıldırım

Yardımcı Tez Danışmanı: Doç. Dr. Tuğba Taşkaya Temizel

Eylül 2021, 188 Sayfa

Video tabanlı öğrenme, kitlesel açık çevrimiçi derslerin (MOOCs) sayısının artmasıyla beraber çok popüler hale gelmiştir. Videoyla etkileşime giren öğrenciler, doğal olarak duraklatma, ileri gitme ve geri gitme gibi tıklama etkileşimlerine dayalı izleme davranışları sergilemektedir. Bunlar öğrenme değişkenlerini tahmin etmek ve bir müdahale çerçevesi geliştirmek için önemli girdilerdir. Ancak bu izleme davranışlarının öğrencilerin lehine nasıl kullanılacağına dair bir kılavuz veya yöntem yoktur. Bu çalışmanın amacı öğrencilerin izleme davranışlarını incelemek ve bunları video derslerin etkisini arttırmak amacıyla bir çerçeve geliştirilmesinde kullanmaktır. Bu doğrultuda, üç deney gerçekleştirilmiştir. İlk olarak, 22 öğrenci ile pilot bir çalışma yürütülmüştür. İkinci olarak, çerçeveyi geliştirmek için 29 öğrenci ile daha kapsamlı bir deney gerçekleştirilmiştir. Üçüncü olarak, çerçeveyi değerlendirmek için 18 öğrenci ile tek grup öntest-sontest modeline göre tasarlanmış bir deney gerçekleştirilmiştir. Bu çalışmalar hem çerçeve hem de kapsamlı öğrenme analitikleri açısından önemli sonuçlar ortaya koymaktadır. Çerçeve ile ilgili sonuçlar, motivasyon kontrol edildiğinde çerçeveye göre oluşturulan etkileşimli videolarda öğrenci memnuniyetinin önemli ölçüde arttığını göstermektedir. Ek olarak, etkileşimli videolarda öğrencilerin önemli noktalara geri dönme sıklığı önemli ölçüde azalmaktadır ve öğrencilerin önemsiz noktaları atlama sıklığı önemli ölçüde artmaktadır. Çerçevenin dışında, sonuçlar; ortalama geriye gidiş hızlarının, duraklama sayısının ve geri gitme sayısının/yüzdesinin öğrencilerin notlarını tahmin etmede en önemli göstergeler olduğunu göstermektedir. Ayrıca sonuçlar video tıklama davranışlarının notu tahmin edebildiğini, memnuniyeti başarılı bir şekilde tahmin etmek için davranışların arkasındaki nedenlerin de gerekli olduğunu göstermektedir. Ancak video tıklama davranışları ve nedenleri etkin katılımı anlamlı bir şekilde tahmin etmek için yeterli değildir.

**Anahtar Kelimeler:** Müdahale Çerçevesi, Öğrenme Analitiği, Eğitsel Veri Madenciliği, Video Tabanlı Öğrenme, Etkileşimli Videolar

*To My Beloved Wife*

## ACKNOWLEDGEMENTS

First of all, I would like to thank my supervisor, Prof. Dr. Soner Yıldırım, for his endless support, motivation, guidance, and encouragement throughout all phases of this dissertation. I am grateful to him for believing in me from the very beginning, and that I would successfully complete my Ph.D. With my deepest gratitude, I also want to thank my co-supervisor, Assoc. Prof. Dr. Tuğba Taşkaya Temizel, for sharing her valuable knowledge and experience. I always progressed confidently in my dissertation from the beginning to the end thanks to her enlightening and creative ideas. I feel very lucky to have worked with these two wonderful academicians, one of whom is an expert in the field of education and the other in the field of data mining.

I also want to thank my thesis supervising committee members, Prof. Dr. Halil Yurdugül and Prof. Dr. Sevgi Özkan Yıldırım, in helping me to improve my dissertation and always ensuring that I was moving in the right direction. I am also thankful to my thesis examining committee members, Assoc. Prof. Dr. Altan Koçyiğit and Assoc. Prof. Dr. Evren Şumuer, for their invaluable feedback and comments.

I would also like to thank Enocta Company, especially Burak Yücelyiğit, Süleyman Cincioğlu, and Okay Yıldırım, for helping to integrate video analytics features into the learning environment. I am also thankful to Merve Aytekin, a dear colleague and friend, for her support during the data collection process, and to Berkan Çelik for his comments during that stage time. I would also like to thank Didem Tufan for her suggestions at the beginning of this dissertation process.

I am also thankful to my dear friends Selcan Kilis and Engin Demirtaş for being there for me whenever I needed ideas or suggestions.

I am also deeply grateful to my two lovely cats, Asya and Suzi, for their becoming a source of joy and peace, and especially for not leaving me alone during the writing of this dissertation.

Last, but surely not least, I would like to express my sincere gratitude to my beloved wife Gülşah for her endless support, understanding, and patience. You were there for me throughout all the difficulties of this long journey. Thank you my darling for being who you are.

**TABLE OF CONTENTS**

ABSTRACT ..... iv
ÖZ ..... v
ACKNOWLEDGEMENTS ..... vii
TABLE OF CONTENTS ..... viii
LIST OF TABLES ..... xi
LIST OF FIGURES ..... xv
LIST OF ABBREVIATIONS ..... xvii
CHAPTERS
INTRODUCTION ..... 1
1.1 Background of the Study ..... 1
1.2 Statement of the Problem ..... 2
1.3 Purpose of the Study ..... 3
1.4 Significance of the Study ..... 4
1.5 Research Questions ..... 4
1.6 Research Strategy ..... 5
1.7 Organization of the Thesis ..... 6
LITERATURE REVIEW ..... 9
2.1 Video-Based Learning ..... 9
2.1.1 Video Types ..... 10
2.1.2 Teaching Methods for Video-Based Learning ..... 10
2.2 Video-Based Learning Analytics ..... 11
2.2.1 Video Viewing Behaviors ..... 12
2.2.2 Clickstream Interactions Based on Video Viewing Behaviors ..... 12
2.2.3 Data Mining Approaches ..... 13
2.2.4 Tools for Video-Based Learning Analytics ..... 16
2.2.5 Predictions ..... 18

2.3 Learning Analytics Interventions ..... 20
2.4 Interactive Videos ..... 21
2.5 Learning Analytics Framework ..... 22
2.6 Learning Variables in Online Courses ..... 23
2.6.1 Satisfaction ..... 23
2.6.2 Engagement ..... 24
2.6.3 Learning Performance ..... 25
RESEARCH METHOD ..... 27
3.1 Research Design ..... 27
3.2 Experiment 1 (Pilot Study) ..... 28
3.2.1 Participants ..... 29
3.2.2 Data Collection Instruments ..... 29
3.2.3 Procedure ..... 30
3.2.4 Results ..... 31
3.3. Experiment 2 ..... 33
3.3.1 Participants ..... 33
3.3.2 Data Collection Instruments ..... 33
3.3.3 Procedure ..... 34
3.3.4 Results (Preliminary Evaluation of Framework) ..... 39
3.4 Experiment 3 (Main Study) ..... 48
3.4.1 Participants ..... 48
3.4.2 Data Collection Instruments ..... 48
3.4.3 Procedure ..... 50
3.5 Data Analysis ..... 51
RESULTS ..... 55
4.1 Results for Framework ..... 55
4.1.1 Differences with Motivation Not Controlled ..... 55
4.1.2 Differences with Motivation Controlled ..... 57
4.1.3 Differences Based on Click Reasons ..... 63
4.2 Results for Comprehensive Learning Analytics ..... 68

4.2.1 Correlations Between Video Clickstream Behaviors and Learning Performance ..... 70
4.2.2 Clusters Based on Video Clickstream Behaviors ..... 72
4.2.3 Predictions Based on Video Viewing Behaviors ..... 82
4.3 Threats to Validity ..... 127
CONCLUSION ..... 129
5.1 Implications for Learning Design and Learning Experience Design ..... 132
5.2 Limitations and Future Work ..... 134
REFERENCES ..... 137
APPENDICES ..... 155
APPENDIX A ..... 155
APPENDIX B ..... 161
APPENDIX C ..... 165
APPENDIX D ..... 167
APPENDIX E ..... 169
APPENDIX F ..... 171
APPENDIX G ..... 173
APPENDIX H ..... 177
APPENDIX I ..... 179
APPENDIX J ..... 181
APPENDIX K ..... 183
CURRICULUM VITAE ..... 185

## LIST OF TABLES

**Table 1.** Research Strategy ..... 6
**Table 2.** Example Features Used in the Literature ..... 13
**Table 3.** Video Lectures in Pilot Study ..... 30
**Table 4.** Number of Interactions ..... 35
**Table 5.** Reasons of Watching Behaviors ..... 36
**Table 6.** Number of Interventions ..... 39
**Table 7.** Survey Results for Each Intervention (Video Lecture 1) ..... 40
**Table 8.** Results for Each Type of Intervention (Video Lecture 1) ..... 41
**Table 9.** Results for Each Intervention (Video Lecture 2) ..... 42
**Table 10.** Results for Each Type of Intervention (Video Lecture 2) ..... 43
**Table 11.** Results for Each Intervention (Video Lecture 3) ..... 44
**Table 12.** Results for Each Type of Intervention (Video Lecture 3) ..... 44
**Table 13.** Comparison of Mean Scores for Satisfaction in Experiment 2 (Video Lecture 1) ..... 45
**Table 14.** Comparison of Mean Scores for Engagement in Experiment 2 (Video Lecture 1) ..... 46
**Table 15.** Comparison of Mean Scores for Satisfaction in Experiment 2 (Video Lecture 2) ..... 46
**Table 16.** Comparison of Mean Scores for Engagement in Experiment 2 (Video Lecture 2) ..... 47
**Table 17.** Comparison of Mean Scores for Satisfaction in Experiment 2 (Video Lecture 3) ..... 47
**Table 18.** Comparison of Mean Scores for Engagement in Experiment 2 (Video Lecture 3) ..... 47
**Table 19.** Linear Video Lectures in Main Study ..... 49
**Table 20.** Interactive Video Lectures in Main Study ..... 49
**Table 21.** Comparison of Mean Scores for Satisfaction ..... 56
**Table 22.** Comparison of Mean Scores for Engagement ..... 56
**Table 23.** Comparison of Mean Scores for Learning Performance ..... 57
**Table 24.** Comparison of Mean Scores for Satisfaction (when motivation was controlled) ..... 59
**Table 25.** Comparison of Mean Scores for Engagement (when motivation was controlled) ..... 60
**Table 26.** Comparison of Mean Scores for Learning performance (when motivation was controlled) ..... 61
**Table 27.** Summary Findings for RQ1 and RQ2 ..... 62
**Table 28.** Comparison of Notetaking between Linear and Interactive Videos ..... 63
**Table 29.** Comparison of Losing Concentration between Linear and Interactive Videos ..... 64

**Table 30.** Comparison of Not Understanding Content between Linear and Interactive Videos ..... 64
**Table 31.** Comparison of Going Back to Review Important Points between Linear and Interactive Videos ..... 65
**Table 32.** Comparison of Skipping Unimportant Points between Linear and Interactive Videos ..... 65
**Table 33.** Comparison of Wanting to Finish Quickly between Linear and Interactive Videos ..... 66
**Table 34.** Comparison of Facing Language Issues between Linear and Interactive Videos ..... 66
**Table 35.** Comparison of Reasons between Linear and Interactive Videos ..... 67
**Table 36.** Number of Interactions for Participants per Video Lecture ..... 69
**Table 37.** Summary of Correlation Findings ..... 72
**Table 38.** Clusters Based on Numbers of Pauses, Forwards and Backwards ..... 74
**Table 39.** Robust Linear Mixed-Model Results to Predict Learning Performance for Clusters Based on Numbers ..... 75
**Table 40.** Robust Linear Mixed Model Results to Predict Learning Performance for Clusters Based on Numbers ..... 76
**Table 41.** Clusters Based on Percentage of Pauses, Forwards and Backwards ..... 77
**Table 42.** Robust Linear Mixed Model Results to Predict Learning Performance for Clusters Based on Percentage ..... 78
**Table 43.** Clusters Based on Forward Speed and Backward Speed ..... 80
**Table 44.** Robust Linear Mixed Model Results to Predict Learning Performance for Clusters Based on Speed ..... 81
**Table 45.** Summary of Clustering Findings ..... 82
**Table 46.** Main Prediction Stages ..... 83
**Table 47.** Features Based on Video Clickstream Behaviors ..... 84
**Table 48.** Features Based on Click Reasons ..... 84
**Table 49.** Hyperparameter Tuning (Parameters and Values) for XGBoost ..... 86
**Table 50.** Importance of Features with XGBoost for Learning performance ..... 86
**Table 51.** Comparison of Three Features and Eight Features with XGBoost for Learning Performance ..... 87
**Table 52.** Importance of Selected Features with XGBoost for Learning Performance ..... 87
**Table 53.** Significance of Features with Robust LMM for Learning Performance ..... 88
**Table 54.** Comparison of Three Features with Eight Features with Robust LMM for Learning Performance ..... 89
**Table 55.** Fixed Effects of Final Robust LMM for Learning Performance ..... 89
**Table 56.** Comparison of XGBoost, Robust LMM and Baseline for Learning Performance Prediction (Random Split) ..... 90
**Table 57.** Comparison of XGBoost, Robust LMM and Baseline for Learning Performance Prediction (One Person Leave Out) ..... 92

**Table 58.** Comparison of XGBoost, Robust LMM and Baseline for Learning Performance Prediction (One Video Leave Out) ..... 94
**Table 59.** Importance of Features with XGBoost for Satisfaction ..... 97
**Table 60.** Comparison of Three Features and Eight Features with XGBoost for Satisfaction ..... 97
**Table 61.** Importance of Selected Features with XGBoost for Satisfaction ..... 98
**Table 62.** Significance of Features with Robust LMM for Satisfaction (Stage 1) ..... 98
**Table 63.** Comparison of Three Features and Eight Features with Robust LMM for Satisfaction (Stage 1) ..... 99
**Table 64.** Fixed Effects of Final Robust LMM for Satisfaction (Stage 1) ..... 99
**Table 65.** Comparison of XGBoost, Robust LMM and Baseline for Satisfaction Prediction (Random Split in Stage 1) ..... 99
**Table 66.** Comparison of XGBoost, Robust LMM and Baseline for Satisfaction Prediction (One Person Leave Out in Stage 1) ..... 101
**Table 67.** Comparison of XGBoost, Robust LMM and Baseline for Satisfaction Prediction (One Video Leave Out in Stage 1) ..... 102
**Table 68.** Significance of Features with Robust LMM for Satisfaction (Stage 2) ..... 105
**Table 69.** Comparison of Nine Features and 15 Features with Robust LMM for Satisfaction (Stage 2) ..... 106
**Table 70.** Fixed Effects of Final Robust LMM for Satisfaction (Stage 2) ..... 106
**Table 71.** Comparison of Robust LMM and Baseline for Satisfaction Prediction (Random Split in Stage 2) ..... 107
**Table 72.** Comparison of Robust LMM and Baseline for Satisfaction Prediction (One Person Leave Out in Stage 2) ..... 108
**Table 73.** Comparison of Robust LMM and Baseline for Satisfaction Prediction (One Video Leave Out in Stage 2) ..... 109
**Table 74.** Importance of Features with XGBoost for Engagement ..... 112
**Table 75.** Comparison of Two Features and Eight Features with XGBoost for Engagement ..... 112
**Table 76.** Importance of Selected Features with XGBoost for Engagement ..... 113
**Table 77.** Significance of Features with Robust LMM for Engagement (Stage 1) ..... 113
**Table 78.** Comparison of One Feature and Eight Features with Robust LMM for Engagement (Stage 1) ..... 114
**Table 79.** Fixed Effects of Final Robust LMM for Engagement (Stage 1) ..... 114
**Table 80.** Comparison of XGBoost, Robust LMM and Baseline for Engagement Prediction (Random Split in Stage 1) ..... 115
**Table 81.** Comparison of XGBoost, Robust LMM, and Baseline for Engagement Prediction (One Person Leave Out in Stage 1) ..... 116
**Table 82.** Comparison of XGBoost, Robust LMM and Baseline for Engagement Prediction (One Video Leave Out in Stage 1) ..... 117
**Table 83.** Significance of Features with Robust LMM for Engagement (Stage 2) ..... 120

**Table 84.** Comparison of Two Features and 15 features with Robust LMM for Engagement (Stage 2) ..... 120
**Table 85.** Fixed Effects of Final Robust LMM for Engagement (Stage 2) ..... 121
**Table 86.** Comparison of XGBoost, Robust LMM and Baseline for Engagement Prediction (Random Split in Stage 2)..... 121
**Table 87.** Comparison of XGBoost, Robust LMM and Baseline for Engagement Prediction (One Person Leave Out in Stage 2)..... 122
**Table 88.** Comparison of XGBoost, Robust LMM and Baseline for Engagement Prediction (One Video Leave Out in Stage 2)..... 124
**Table 89.** Summary of Prediction Findings..... 126

## LIST OF FIGURES

**Figure 1.** Research Design ..... 28
**Figure 2.** Pilot Study Steps ..... 30
**Figure 3.** Pearson Correlation Matrix ..... 32
**Figure 4.** Framework (First Version) ..... 37
**Figure 5.** Framework (Final Version) ..... 38
**Figure 6.** BIC Values for Number of Clusters Based on Number of Clicks ..... 73
**Figure 7.** Cluster Quality Based on Numbers of Clicks ..... 74
**Figure 8.** Cluster Distribution Based on Numbers of Pauses, Forwards, Backwards ..... 75
**Figure 9.** Cluster Distribution Based on Percentage of Pauses, Forwards, Backwards ..... 78
**Figure 10.** BIC Values for Number of Clusters Based on Click Speed ..... 79
**Figure 11.** Cluster Quality (Based on Click Speed) ..... 80
**Figure 12.** Cluster Distribution Based on Forward Speed and Backward Speed ..... 81
**Figure 13.** Learning Performance Predictions (based on Random Split) for 42 Observations ..... 91
**Figure 14.** Learning Performance Predictions (based on One Person Leave Out) for 42 Observations ..... 93
**Figure 15.** Learning Performance Predictions (based on One Video Leave Out) for 42 Observations ..... 95
**Figure 16.** Satisfaction Predictions (based on Random Split) for 42 Observations in Stage 1 ..... 100
**Figure 17.** Satisfaction Predictions (based on One Person Leave Out) for 42 Observations in Stage 1 ..... 101
**Figure 18.** Satisfaction Predictions (based on One Video Leave Out) for 42 Observations in Stage 1 ..... 103
**Figure 19.** Satisfaction Predictions (based on Random Split) for 42 Observations in Stage 2 ..... 107
**Figure 20.** Satisfaction Predictions (based on One Person Leave Out) for 42 Observations in Stage 2 ..... 109
**Figure 21.** Satisfaction Predictions (based on One Video Leave Out) for 42 Observations in Stage 2 ..... 110
**Figure 22.** Engagement Predictions (based on Random Split) for 42 Observations in Stage 1 ..... 115
**Figure 23.** Engagement Predictions (based on One Person Leave Out) for 42 Observations in Stage 1 ..... 117
**Figure 24.** Engagement Predictions (based on One Video Leave Out) for 42 Observations in Stage 1 ..... 118
**Figure 25.** Engagement Predictions (based on Random Split) for 42 Observations in Stage 2 ..... 122

**Figure 26.** Engagement Predictions (based on One Person Leave Out) for 42 Observations in Stage 2 ..... 123
**Figure 27.** Engagement Predictions (based on One Video Leave Out) for 42 Observations in Stage 2 ..... 124

## LIST OF ABBREVIATIONS

| | |
|---|---|
| **ANN** | Artificial Neural Network |
| **BIC** | Schwarz's Bayesian Criterion |
| **CFA** | Correctness on First Attempt |
| **CEIT** | Computer Education and Instructional Technology |
| **CSV** | Comma-Separated Values |
| **KNN** | K-nearest Neighbor |
| **LAK** | Conference on Learning Analytics and Knowledge |
| **LMM** | Linear Mixed Model |
| **LMS** | Learning Management System |
| **LSS** | Learner Satisfaction Scale |
| **MAE** | Mean Absolute Error |
| **METU** | Middle East Technical University |
| **MOOCs** | Massive Open Online Courses |
| **RMSE** | Root Mean Square Error |
| **RQ** | Research Question |
| **SD** | Standard Deviation |
| **SIMS** | The Situational Motivation Scale |
| **SVM** | Support Vector Machines |
| **VES** | Video Engagement Scale |
| **VIF** | Variance Inflation Factor |
| **WaVe** | Workshop on Analytics on Video-Based Learning |

# CHAPTER 1

# INTRODUCTION

This chapter presents the background of the study, the statement of the problem, purpose of the study, significance of the study, the research questions, and the research strategy in order to introduce the research in detail.

## 1.1 Background of the Study

Video-based learning has attracted the interest of many instructors and researchers with the information and communication technology developments seen in the last decade, even though its roots were established many years previously. The main reasons behind this attraction are today's high network bandwidth and the diversity and availability of portable computing devices (Ronchetti, 2013) (e.g., laptop, tablet computers, smartphones, etc.). The interest in video-sharing platforms such as YouTube and Vimeo has also played an important role in the rapid acceptance of video-based learning (Adnan & Redzuan, 2017). In addition, studies explaining the benefits of video-based learning have also encouraged instructors to adopt video-based learning (Giannakos et al., 2014). Instructors face an ever-increasing need to disseminate information to large student groups due to the increasing numbers of course participants, which has led to the steadily increasing availability of Massive Online Open Courses (MOOCs). High numbers of courses and registered learners in the most known MOOCs such as Coursera, EdX, Udemy, and Khan Academy clearly demonstrate the popularity of MOOCs, through which many users are able to reach video lectures from wherever and whenever they want.

During video lectures, students perform video viewing behaviors based on clickstream interactions, such as *pause*, *backward* and *forward* (Shi et al., 2015). These viewing behaviors have attracted the attention of researchers and educators interested in the areas of educational data mining and learning analytics. Their focus has started to consider how video viewing behaviors can be applied in order to improve student learning. This interest has led to a new field of study known as "video-based learning analytics" or in short, "video analytics" (Giannakos et al., 2013; Mirriahi & Vigentini, 2017). The aim of video

analytics is to increase the effect of video lectures on students' learning based on their video watching behavior data.

Research conducted by Columbia University showed that the completion rate of courses shared on edX and Coursera was 15% or less (Hollands & Kazi, 2018). In addition, in a similar study conducted by the Massachusetts Institute of Technology (MIT) and Harvard University, the dropout rate was found to be more than 90% (Ho et al., 2014), and this rate had not been found to improve over the subsequent 6 years (Reich & Ruipérez-Valiente, 2019). The main reason attributed to the high level of dropout is a lack of interactivity in video-based learning platforms (Chatti et al., 2016). Interactivity can be provided with real time in-video interventions. By analyzing students' video viewing behaviors, intervention points can be extracted (Akçapınar & Bayazıt, 2018) in order to show the best timing for interventions based on considering the peak points at which students' interactions rise (Kim, Li, et al., 2014). The interventions aim to direct students' watching behaviors to desired behaviors so that the positive effect of video lectures on students' learning can increase.

In addition, students' video viewing behaviors can provide good indicators to determine students at risk in terms of learning performance, satisfaction, or level of engagement. These three variables are considered essential metrics in evaluating the effectiveness of video lectures. Learning performance carries vital importance in measuring student understanding of a specific subject, whilst engagement related to their active participation, and satisfaction pertains to the fulfillment of student expectations and needs regarding course materials and content. Predictors of learning performance, engagement, and satisfaction demonstrate which students require some form of intervention. Since students create large amounts of data from their video viewing behaviors when watching video lectures, determining predictors in terms of these viewing behaviors is deemed useful and thereby highly applicable for platforms offering video-based learning.

### 1.2 Statement of the Problem

Video viewing behaviors based on clickstream interactions are seen as important indicators to determine students at risk (Mubarak, 2020). If students exhibit low satisfaction or low engagement levels during their video lecture watching and/or subsequently perform subpar in their exams, this may indicate potentially ineffective video lectures having been viewed. Therefore, determining students at risk during video lectures can help instructors to take prompt action in order to modify their students' behaviors so that increased benefit can be gained from their video lecture watching. However, even with numerous studies in the literature on predicting learning performance, satisfaction, and engagement, limited studies exist which focus on the video viewing

behaviors of learners using video-based learning platforms to predict students' learning performance, satisfaction, and engagement.

Linear video lectures might also result in students dropping courses (Chatti et al., 2016); hence one reason many MOOCs are still far from being considered as interactive forms of teaching. Since videos form the main component of MOOCs or video sharing platforms, making videos more interactive will naturally meet the requirement to improve interactivity in MOOCs. It is suggested, therefore, that timely and successful interventions added to video lectures will both improve the interactivity of MOOCs and their effectiveness for learners (Rienties, Boroowa, et al., 2016). Video viewing behaviors can be used to identify intervention points in video lectures so that interactive elements can be effectively added to these points (Kim, Li, et al., 2014). However, educators are still largely unaware of how to make best use of video viewing data pertaining to their students' viewing behaviors, and to make linear videos more interactive in favor of their students.

This background demonstrates a need to develop a framework based on video viewing behaviors that can be applied in order to increase the impact and effectiveness of video lectures, and to investigate the impact of video viewing behaviors on students' learning performance, engagement, and satisfaction.

### 1.3 Purpose of the Study

The purpose of this study is to investigate learners' video viewing behaviors through the application of suitable data mining approaches, and then to develop a framework for the design and development of interactive video lectures based on students' video viewing behaviors in order to increase the impact and effectiveness of video lectures. As such, there are two important outputs of the study:

(1) Framework Based on Video Viewing Behaviors
(2) Predictors of Learning Performance, Satisfaction, and Engagement

The framework will show how click types are transformed into interactive elements with five questions ("where, why, which, how, what?"). As a result, it will be possible to add interactive elements meaningfully to linear videos based on students' video viewing behaviors.

Predictors will demonstrate the impact of video clickstream behaviors on students' learning performance, engagement, and satisfaction. The study will also show correlations, clusters together with predictors of students' learning performance, engagement, and satisfaction in order to see a more detailed effect of each different clickstream behavior. Based on this data-driven knowledge, it is aimed that instructors

may be better able to determine students at risk in terms of their learning performance, engagement, and satisfaction.

### 1.4 Significance of the Study

There are certain important contributions expected from this study:

First, the study will produce a framework based on learners' video viewing behaviors. The framework can then be used to guide instructors about how to create more interactive videos based on students' video viewing behaviors and to help their students to learn in more interactive and enjoyable ways. The framework is also intended to be helpful in improving existing courses, especially those published on MOOC platforms. Since the framework is also intended to be a guideline for changing existing linear to interactive videos, it is expected to be a cost effective course of action when compared to the creation of interactive videos from scratch based on existing linear videos. The framework may also form a basis for the formulation of intelligent and adaptive e-learning systems.

Second, the study aims to enable educators and researchers to determine those students who are at risk by predicting their learning performance, engagement, and/or satisfaction based on students' video viewing behaviors. That is, the study aims to provide significant information regarding the video viewing behaviors of students, and to present indicators of both positive and negative behaviors in terms of students' learning performance, engagement, and satisfaction.

Third, the study aims to contribute significantly to the literature in terms of video-based learning analytics; demonstrating the importance of video analytics in a practical application. The study also aims to demonstrate which data is considered useful, and how it should be analyzed in order to improve the efficiency of video lectures, and thereby reveal important insights for both researchers and practitioners.

Finally, and most significantly, learners will be able to spend their time more effectively during video lecture watching based on the help this study aims to provide.

To summarize, this study is expected to be considered beneficial for three sets of stakeholders: learners, instructors, and researchers, as the most critical actors in the whole teaching and learning process.

### 1.5 Research Questions

There are six research questions in the study. The first three questions concern the framework that will be developed based on learners' video viewing behaviors, while the

last three questions focus on the correlations, clusters, and predictors based on students' viewing behaviors.

> RQ1: Are there any significant differences between linear and interactive videos in terms of students' satisfaction, engagement, and learning performance?
>
> RQ2: Are there any significant differences between linear and interactive videos in terms of students' satisfaction, engagement, and learning performance when motivation is controlled?
>
> RQ3: Are there any significant differences between linear and interactive videos in terms of students' click reasons?
>
> RQ4: Are there any correlations between video clickstream behaviors and learning performance?
>
> RQ5: Are there any significant differences between clusters within cluster sets based on video clickstream behaviors in terms of learning performance?
>
> RQ6: Which video viewing behaviors and to what extent can these behaviors predict students' engagement, satisfaction and learning performance?

RQ1 and RQ2 investigate the effect of the developed framework based on students' satisfaction, video engagement, and learning performance scores, both with and without considering student motivation. In addition, RQ3 analyzes which behaviors change according to the application of the developed framework. RQ4 investigates correlations between video clickstream behaviors and learning performance, while RQ5 studies clusters based on students' video clickstream behaviors and differences among clusters in terms of learning performance. Finally, RQ6 aims to predict learner satisfaction, engagement, and learning performance scores according to their video viewing behaviors (video clickstream behaviors and click reasons), and to identify their predictors.

### 1.6 Research Strategy

Three different studies are performed in the study in order to answer the research questions. Table 1 introduces each study, including its research design, aim, and participants. Each are then presented in detail in the Research Method section.

**Table 1.** Research Strategy

| Study | Research Design | Aim | Participants |
|---|---|---|---|
| **Experiment 1 (Pilot Study)** | One shot case study design | Assess the capability of current video analytic tools.<br>Identify correlations between students' video clickstream behaviors and learning performance.<br>Help in designing Experiment 2. | 22 students enrolled to "Principles and Methods of Instruction" course (2017-2018 Spring semester) |
| **Experiment 2** | One group pretest-posttest design | Develop a framework based on video viewing behaviors. | 29 students enrolled to "Principles and Methods of Instruction" course (2018-2019 Spring semester) |
| **Experiment 3 (Main Study)** | One group pretest-posttest design | Evaluate the developed framework.<br>Investigate video viewing behaviors' impact on students' learning performance, engagement, and satisfaction by focusing on correlations, clusters, and predictors. | *18/16 students enrolled to "Principles and Methods of Instruction" course (2019-2020 Fall semester) |

*18/16 relates to 18 students used to evaluate the developed framework, while data for 16 students were used to investigate their video viewing behaviors through comprehensive learning analytics. Details are explained in Section 4.2.

## 1.7 Organization of the Thesis

This thesis is presented as five chapters and is organized as follows:

Chapter 2 presents the related works from the literature regarding video-based learning, video-based learning analytics, interventions, interactive videos, and learning analytics framework.

Chapter 3 presents the research methodology including the research design. In addition, the participants, data collection instruments, and procedures are explained in detail for each of the study's three experiments.

Chapter 4 presents the results of the study through association with the research questions for both the framework and clickstream behaviors.

Finally, Chapter 5 discusses and presents a conclusion to the study's findings. Also included in this chapter are implications for learning design and learning experience design, as well as limitations of the current study and an outlook to future works.

# CHAPTER 2

# LITERATURE REVIEW

This chapter presents the related works in the literature that have researched video-based learning, video-based learning analytics, interventions, interactive videos, and learning analytics framework. In the section on Video-Based Learning, the history and advantages of video-based learning, video types, and the teaching methods used in video-based learning are detailed. In the Video-Based Learning Analytics section, the published literature on watching behaviors, clickstream interactions, data mining approaches, tools for video-based learning, and predictions are presented. In the section on Learning Analytics Interventions, previous studies about learning analytics interventions are mentioned, whilst in the Interactive Videos section, detailed information is provided about the literature on interactive videos and interactivity techniques used in videos. Finally, in the section on Learning Analytics Framework, existing frameworks that have been published are introduced.

## 2.1 Video-Based Learning

Video-based learning is a popular learning approach today, with increasing numbers of MOOCs and video-sharing platforms such as YouTube and Vimeo used in the educational setting. Educational films and television programs, as well as video cassettes and DVDs have also historically contributed to the development of today's video-based learning (Kleftodimos & Evangelidis, 2016a; Yousef et al., 2014). Increases in internet/network bandwidth and the variety of devices now commonly available has facilitated access to video lectures for the masses (Ronchetti, 2013). This has substantially increased the popularity of video-based learning across all areas. Lastly, the COVID-19 pandemic has forced many to access lectures online from their home in the absence of traditional face-to-face lectures and classroom/laboratory teaching venues. Video-based learning has now become an essential part of today's education.

In addition, studies have described various benefits of video-based learning in order to encourage teachers to utilize video-based teaching within their lessons (Giannakos et al., 2014). Considering the various published studies in this area, the most important and

significant benefits of video lectures can be summarized as follows (Yürüm & Yıldırım, 2020):

- Video lectures can be accessed with relative ease from almost any device;
- Video lectures reduce learning costs as they can be reached by many users at the same time, and can also be reused; and,
- Video lectures can be accessed at virtually any time and any place.

### 2.1.1 Video Types

There are different types of videos employed in education. In the literature, the types of videos are determined according to the purpose of the study. Decomposition criteria can be production style, interactivity, or time as a means of categorizing videos into specific types. For example, some studies categorize videos according to production styles such as Khan Style, talking head, PowerPoint slides with a voice, etc. (Kim, Li, et al., 2014; Sinha et al., 2014). On the other hand, others base their categorization on videos being either interactive or linear (Kleftodimos & Evangelidis, 2014b), and Mirriahi and Vigentini classified videos as being either synchronous or asynchronous based on time as a deciding factor (2017).

### 2.1.2 Teaching Methods for Video-Based Learning

The growing interest in video-based learning has led to the emergence of a significant number of teaching methods that include the use of video content as a teaching medium. According to Yousef et al. (2014), these teaching methods can be categorized into six areas, with collaborative learning, micro teaching, video summarization, video assessment, hybrid learning, and student-centered learning. The collaborative learning approach expects students to work together, sharing their ideas, resources, and activities during or after watching video content. While video summarization methods aim to help students learn by summarizing the video content they have seen, video assessment methods aim to help teachers evaluate student progress. Micro teaching is a method that uses the medium of video in teacher evaluation and assessment. In addition, hybrid teaching methods have also become popular today, with the flipped classroom being one such method that has attracted many educators because it enables learners to learn more subjects within a shorter period. Other than these methods, some published studies have utilized videos in student-centered learning. However, Yousef et al. (2014) emphasized that only 15% of studies relate to student-centered learning, while 85% are teacher-centered. Today, MOOCs are at the heart of video-based learning; therefore, when the large number of learners enrolled in MOOCs is considered, the potential for turning this area to being student-centered presents some considerable barriers for educators.

All of these teaching methods show the popularity of video-based learning, even though it may not be easy to implement many such methods within MOOCs. In addition, these teaching methods are also examples of the academic application of videos.

## 2.2 Video-Based Learning Analytics

Learning analytics aims to improve the efficiency of education (Volungevičienė et al., 2019). Naturally, video-based learning analytics aims to enhance learning with educational data obtained from platforms that include video lectures (Giannakos et al., 2016). Therefore, video-based learning analytics, sometimes referred to as video analytics in educational articles, has attracted researchers as a means to improving student learning via data-driven approaches.

Mirriahi and Vigentini defined video-based learning analytics as "the collection, measurement, and analysis of such data for the purposes of understanding how learners use videos" (2017, p. 251) and in integrating videos into the definition of learning analytics put forwards by Siemens (2013). Studies in this area have increased steadily following on from the initial workshop held on analytics from video-based learning (WaVe 2013), which was organized as part of the 3rd Conference on Learning Analytics and Knowledge (LAK 2013).

Video-based learning analytics is seen as a useful means to understanding the effectiveness and design essentials of video lectures (Akçapınar & Bayazıt, 2018). In addition, it is also helpful for use in understanding students' learning performance and also to improve teaching techniques (Giannakos et al., 2015). With the help of analytics, educators are better able to appraise video lectures in the flipped classroom or MOOC setting in terms of their differing perspectives (Lau et al., 2018).

In this context, analytical approaches for video-based learning provide benefits for four different stakeholder groups (students, teachers/instructors, researchers, and administrators) (Yürüm & Yıldırım, 2020), therefore:

- Students will be able to follow video lessons according to their personal needs;
- Teachers will be able to access important tips on the improvement of video lectures;
- Researchers will be able to investigate variables such as dropout, engagement, and learning performance by collecting and utilizing certain metrics in video lectures; and,
- Administrators will be better able to make quick decisions regarding video lectures.

### 2.2.1 Video Viewing Behaviors

Researchers investigate learners' video viewing behaviors in order to understand their video viewing habits (Hildebrand & Ahn, 2018). While some behaviors relate to general properties such as video popularity and watching time (Qiu & Cui, 2010; Zhou et al., 2014), others focus on specific click types such as *pause*, *forward*, and *backward* (Conole et al., 2015; Lemay & Doleck, 2020; Sinha et al., 2014; Yang et al., 2018; Yu et al., 2019). From the studies that have focused on clickstream interactions, some have centered on transitions or sequences between click types in order to spot similar behaviors in users (Sinha et al., 2014; Yu et al., 2019), whereas others have considered the numbers of clickstream interactions (Lu et al., 2018).

### 2.2.2 Clickstream Interactions Based on Video Viewing Behaviors

Video clickstream interactions offer significant potential to improve the outcome of video-based lectures (Kim, Guo, et al., 2014), through revealing valuable insights about different learner behaviors (Giannakos et al., 2014). Interaction peaks are defined as the points within a video where click operations such as *pause*, *backward*, and *forward* are performed the most, indicating either points of interest or confusion for users (Kim, Li, et al., 2014). Interaction patterns, generated by video clickstream data are then used in order to understand the learners' characteristics and also to understand their learning strategies (Conole et al., 2015; Dimitrova et al., 2017; Kleftodimos & Evangelidis, 2014b; Kuo et al., 2015; Ullrich et al., 2013). Researchers have recently investigated the relationships between video clickstream behaviors and variables such as student engagement to a course/lesson and their academic achievement, and have revealed that a set of clickstream behaviors may be employed in the measurement of these variables (Brinton et al., 2016; Kleftodimos & Evangelidis, 2014b; Lemay & Doleck, 2020). There have also been studies published that have focused on learners' watching behaviors which aim at revealing students' learning styles (Brooks et al., 2013; Kleftodimos & Evangelidis, 2014a); however, the published studies have yet to venture beyond determining learning styles or predicting variables such as learner performance.

There are a variety of interactions used in studies based upon the video watching behaviors of students. Video clickstream interactions have also been used as "events" in some studies. The existing literature generally includes one or a number of behaviors as shown in Table 2. In the literature, these interactions have been tracked by using Open edX, an institution's own platform, or a learning management system. The video clickstream interactions found in the literature are: *Play*, *Pause*, *Backward (Seek/Jump)*, *Forward (Seek/Jump)*, *Speed Up*, *Speed Down*, *Mute*, *Unmute*, *Full Screen On*, *Full Screen Off*, *Show Subtitle*, *Hide Subtitle*, *Show Transcript*, *Hide Transcript*, *Download Video*, *Start/View*, *Stop*, *Replay*, *Load*, *Complete/End*, and *Exit Video*. The existing studies

generally include some of these within their research, although *Play* and *Pause* are the mostly widely used interactions in the current literature. Based on these interactions, different types of features have been used in the studies according to their aim or scope. These features are generally derived by using the terms listed in Table 2:

**Table 2.** Example Features Used in the Literature

| Term | Feature |
|---|---|
| Average | Average Time Spent per Video |
| | Average Number of Repetitions per Video |
| Percentage | Percentage of Completed Videos |
| | Percentage of Started Videos |
| Fraction/Proportion | Fraction Completed |
| | Fraction Played |
| Number | Number of Pauses |
| | Number of Plays |
| | Number of Views |
| Time | Time Spent in Pause |
| | Replayed Video Time |
| Total | Total Time Spent |
| | Total Number of Clicks |

The features derived from video clickstream interactions were mostly used together with the additional features related to exercise, navigation, or forum. A few studies have used them directly and focused only upon the video clickstream behaviors. The features generally show either a count or a time as a value, and are either continuous or ratio variables.

### 2.2.3 Data Mining Approaches

Many data mining approaches have been applied in video-based learning, although they vary according to the purpose of the research and the type of data obtained. Algorithms belonging to four basic categories (classification, regression, clustering, and association rules) found in data mining have been used in studies researching video-based learning.

Regression studies are generally used to predict students' learning performance, whilst classification algorithms are used both to measure the probability of dropping out of video lessons and to assign video lectures or students to a particular class. Clustering algorithms are preferred more to reveal previously unspecified patterns of video lessons or students based on in-video behavior, and especially for the determination of different learning styles. Association rules are algorithms used to determine video lessons watched together.

#### 2.2.3.1 Clustering

The existing studies have generally used clustering algorithms to group students or videos based on similar characteristics according to learners' watching behaviors. Each group includes between two and 10 pattern differentials according to the setting environment and participant. Whilst some are named directly with clicks such as *long pause* and *frequent pause*, others are more generically named based on behavioral characteristics such as *regular watcher*, *engaged watcher*, and *pauser watcher*. In total, 10 different clustering examples were noted regarding students' watching behaviors. Many of these examples aim to reveal the characteristics of engaged students, achieved students, interested students, and self-regulated students based on the specific purpose of the research. In doing so, studies are able to determine the learning strategies of students so that they can better design video lectures to address varying types of students in the future. In addition, such studies are helpful in identifying problematic behaviors, which can be changed to desired learning behaviors through appropriate interventions.

The most frequently used algorithm to cluster learning behavior in the literature has been k-means, since it is the most popular algorithm used for grouping. In a study conducted by Kleftodimos and Evangelidis (2016a), a video-based learning environment for learning image editing techniques was used to collect data about learner's behaviors. In their study, four different clusters were discovered using k-means algorithms. In addition to in-video behaviors, the study considered additional actions such as exercise link clicks, save clicks, and section links to cluster students' learning strategies.

Kleftodimos and Evangelidis (2014b) also used k-means algorithm in order to measure learner engagement and video popularity, using different metrics for each. They grouped learners according to linear or interactive video preferences by taking into account attributes such as percentage of linear videos, percentage of interactive elements, number of days viewing, and number of actions. They also employed metrics such as times started, times abandoned, number of learners, and actions per video to measure video popularity. As a result, they found five learner engagement clusters and two video popularity clusters.

Dimitrova et al. (2017) determined three different styles of engagement by using k-means algorithms to guide researchers about intervention times. According to their study, self-

regulated learners form the ideal learner group, while parochial learners require the highest number of interventions.

Kuo et al. (2015) also used k-means as a clustering algorithm, supporting their results with statistics and other data mining approaches such as decision trees and sequential patterns. Their study showed that the use of course videos affected learning outcomes.

In addition to those studies related to learning videos, video viewing behaviors in a movie trailer were also tracked and grouped according to the k-models clustering algorithm, which is an extension of the k-means algorithm (Mongy, 2007).

Differently from these studies, Conole et al. (2015) used Neural Gas, a neural network-based convex clustering algorithm to find video interaction patterns to reveal relationships between video interactions and perceived video difficulty, video revisiting behaviors, and student performance. Their findings emphasized the importance of changing video behaviors and pause times for intervention and interaction, as well as determining revisiting parts for quick access.

#### 2.2.3.2. Classification

The literature includes only a limited number of studies that utilized classification algorithms to analyze students' video watching behaviors. Kuo et al. (2015) classified videos as on time, early, late, and review according to the access time of students by taking the scheduled period into account. In addition, Ullrich et al. (2013) used k-nearest neighbor as a classification algorithm to investigate whether or not viewing patterns such as rewind or skip forward could predict a set of variables including student score, student pass, teacher rating, course ID, course category, student ID, and teacher ID.

#### 2.2.3.3 Regression

Regression algorithms are used to predict the continuously dependent variable. Accordingly, video-based lessons were the more preferred for predicting students' learning performance. Principally, students' learning performance can be predicted using the multiple linear regression algorithm (Conijn et al., 2018; Almeda et al., 2018; Yang et al., 2018). These studies investigated to what extent the metrics related to general usage such as video views predict students' performance. The literature emphasizes that by discovering metrics that predict students' learning performance, courses can be designed in advance in consideration of these metrics, and that students' learning performance can be increased through interventions as deemed necessary.

#### 2.2.3.4 Association Rule Mining

The literature contains only a very limited number of studies using association rule mining in video-based learning. In one such study, Kleftodimos and Evangelidis (2014b) applied association rule mining algorithm in addition to clustering algorithm, with the aim to identify videos watched together within a single session by using the apriori algorithm. Three consecutive videos which were reportedly not easily understood by students during their laboratory session were found to have been viewed mostly together.

#### 2.2.3.5. Summary

The related studies show that many used clustering algorithms, whilst a few applied classification and association rule mining algorithms in video-based learning. Some studies also supported their results with statistical approaches such as ANOVA and Kruskal-Wallis test. Moreover, in studies related to video-based learning, basic data mining approaches were often used as the method of data analysis. However, different machine learning algorithms could be used to improve the prediction and classification performance of video analytics applications.

### 2.2.4 Tools for Video-Based Learning Analytics

There are a number of different tools that offer an analytics feature for video-based learning, and these can be categorized as either "Academic Video Analytic Tools" or "Other Video Analytic Tools."

#### 2.2.4.1. Academic Video Analytic Tools

Several video analytics tools have been developed for the purpose of tracking the video behaviors of students that watch videos from educational video platforms. Whilst some support the video players, others support the video together with the learning environment such as learning management systems.

The most significant video analytics tool published is SocialSkip, as developed by Chorianopoulos et al. (2014) to track *play*, *pause*, *backward*, and *forward* actions of users in YouTube videos. The tool stores behavioral data with six attributes, which are TesterID, Time, TransactionID, TransactionTime, Transaction, and Skiptime as a CSV. Based on the collected data, the tool provides two different graphs for both *backward* and *forward* analysis in order to show the parts that were skipped as *backward* or *forward* on a timeline.

VisMOOC is a video analytics system developed by Shi et al. (2015) to help instructors and researchers to analyze students' video behaviors in MOOC platforms such as

Coursera. VisMOOC has two important graph features: Event Graph and Seek Graph. Event graph shows the number of different events as *play*, *pause*, *seeked*, *ratechange*, *stalled*, and *error in timeline* for different types of videos, whereas the Seek graph shows *backward* and *forward* density on a timeline. With the help of the VisMOOC tool, it is also possible to attain summary statistics for clickstream data and changes in learning behavior over time.

In addition to SocialSkip and VisMOOC, several other video analytics tools have been mentioned in the literature related to video-based learning analytics. These less commonly used tools are the Active Video Watching (AVW) system (Dimitrova et al., 2017), CourseMapper (Chatti et al., 2016), CLAS (Risko et al., 2013), and LIVE Interaction in Virtual Learning Environments (Wachtler, Khalil, et al., 2016).

#### 2.2.4.2. Other Video Analytic Tools

In addition to the academic video analytic tools, current video cloud platforms also include analytics features that could be employed in others areas, including education. YouTube, which is the most popular video site on the web today, also includes its own video analytics feature known as YouTube Analytics. This platform can provide a variety of reports about YouTube videos, categorized as Watch Time Reports and Interaction Reports. Watch Time Reports provide general information about video watching time, audience retention, demographics, playback locations, traffic sources, devices, and translations, whilst Interaction Reports cover subscribers, likes and dislikes, comments, and sharing. However, these all indicate generic interest about a video; that is, you can obtain limited information such as average watching time, numbers of likes and comments etc. In addition, it is not possible to obtain information about specific users' preferences.

Vimeo is another popular video cloud platform, and which includes four important analytical features. First, it provides statistical information such as number of plays, finishes, likes, and comments at a glance. The second feature shows demographic information such as viewers' locations. The third tracks how videos perform over time, with numbers of plays, likes, and finishes shown against a timeline graph for a specific video. The fourth feature creates and exports custom reports including both statistical information and graphics.

#### 2.2.4.3. Summary

All video analytics tools are valuable to a certain degree based on their specific purpose, with each tracking different metrics (clickstream interactions or metrics) about learners or users. In addition, they can also provide limited video analytics; however, these analytics need to be improved in order to increase the effectiveness of video-based learning. A tool

having the flexibility to support the track of each metric and provide more detailed analytics based on these metrics can elicit a greater video-based learning effect. As such, the metrics discussed in this section are significant indicators for those looking to develop their own video analytics tools.

### 2.2.5 Predictions

#### 2.2.5.1 Learning Performance (Grade) Prediction

Predicting the learning performance of students has long been a focus for academic researchers. By developing prediction models or determining the predictors of learning performance, it is possible to detect students who are at risk of course failure or dropout (Mubarak, 2020). These predictions can therefore also be used to provide insight regarding the timing of interventions (Lu et al., 2018). Numerous studies have focused on predicting students' learning performance (Brinton et al., 2015; Guo et al., 2015; Hasnine et al., 2018; Soni et al., 2018); however, only a few have focused on the watching behaviors of students to predict students' learning performance.

Lu et al. (2018) investigated several features such as out-of-class practice behaviors, homework and quiz scores, as well as certain watching behaviors in order to predict students' academic performance on a blended Calculus course. Their study focused on the number of times students registered certain click types each week as watching behaviors. Their results found seven critical factors including number of backwards per week and number of plays per week as watching behavior which were then used to determine student academic performance.

Yang et al. (2017) used a time series neural network to predict students' learning performance, with past quiz performance data and clickstream data employed together as input to predict average Correctness on First Attempt (CFA) scores on quizzes.

A study conducted by Hasan et al. (2020) predicted higher education students' performances using various features such as student academic information, students' activity, and student video interactions, and then compared data mining algorithms in terms of prediction. Random Forest algorithm gave the best results; however, the separate impact of watching behaviors on academic performance could not be inferred from their study since the previous coursework or exam scores were included in the prediction of the students' academic performance.

Yu et al. (2019) determined seven behavioral patterns based on viewing behavior sequences, and then predicted whether or not a student would pass a course based on the total number of each pattern using different algorithms. Similarly, Brinton et al. (2015)

used sequences to find patterns, and then investigated the relationships between these patterns and CFA in quizzes by attempting to predict CFA from recurring patterns.

There are also other studies that have focused on video watching behaviors in order to predict learning performance (Gardner & Brooks, 2018; Lemay & Doleck, 2020; Mbouzao et al., 2020). However, the current literature is only helpful in predicting learning performance to a certain degree. Currently, researchers use either similar features or aggregated features by disregarding that students may perform different behaviors when watching different videos.

As can be seen from the literature, the current trend of predicting learning performance based on watching behaviors has begun to increase in recent years. The current study, however, aims to present an alternative angle by showing the relationships, clusters, and predictions together in order that collectively they will provide researchers and instructors with greater insight with a more holistic picture. In addition to the *number of clicks*, the current study also includes relative *percentage of clicks*, *forward speed*, and *backward speed* as video viewing behaviors. Repeated measures and analysis suited to repeated measures are considered another strong side of the current study.

#### 2.2.5.2. Engagement Prediction

Engagement is a necessary prerequisite for learning (Guo et al., 2014) and, therefore, understanding learner engagement is vital to achieving pedagogical effectiveness from video lecturing in MOOCs (Hu et al., 2020). Historically low completion rates on MOOCs has motivated researchers to focus on the issue of student engagement (Brinton et al., 2016). Clickstream data can provide important information about how students engage with video-based lectures (Lin et al., 2017). By analyzing students' video watching behaviors, researchers can determine if a student skips over certain important points of a video or replays these points (McGowan & Hanna, 2015). There have been a number of studies published that have attempted to predict student engagement according to video watching behaviors (Mongkhonvanit et al., 2019). Körösi et al. (2018) aimed at predicting course completion and student engagement based on clickstream data by comparing different machine learning algorithms. Their results showed that the most effective algorithms were Random Forest and Bagging. Lan et al. (2017) also focused on how behavioral data can be used to model student engagement, whilst Aguiar et al. (2015) focused on using clickstream data to predict student engagement. Xie et al. (2017) modeled student engagement with explicit features such as video watching time. In addition, the effect of video production variables on student engagement was investigated by Van Der Sluis et al. (2016), whilst Guo et al. (2014) examined how video production decisions have a role in student engagement.

These existing studies have shown that students' video watching behaviors based on clickstream data can have an important role in predicting student engagement. By using these predictions, interventions can be applied to improve courses (Mongkhonvanit et al., 2019). Nevertheless, there has been a lack of studies related to student engagement (Giannakos et al., 2016); hence, further investigation is needed in order to clearly reveal those watching behaviors that may predict student engagement. Mubarak (2020) especially emphasized the importance of investigating implicit features such as *pause*, *forward*, or *backward* rather than explicit features such as *views* or *annotations*.

#### 2.2.5.3. Satisfaction Prediction

Satisfaction is one of the most important variables investigated in order to measure the effectiveness of learning (Zhang, 2005). Many studies in the literature have focused on student satisfaction, and these mostly relate to video lectures where student satisfaction is examined by either comparing online learning with traditional learning, or by comparing interactive videos with linear videos (Allen et al., 2002; Choe et al., 2019; Hilgenberg & Tolone, 2000; Roach & Lemasters, 2006; Zhang et al., 2006).

Even where studies have concentrated on satisfaction, the number trying to predict satisfaction has been fewer than for engagement and learning performance. The prediction of satisfaction in MOOCs or in distance learning has usually been examined by focusing on metrics pointing to separate scores for content, structure, assessment, or instructors (DeBourgh, 2003; Hew et al., 2020), whilst satisfaction predictors have yet to be investigated in terms of video viewing behaviors. The current study, therefore, was based on the premise of viewing behaviors being useful in predicting student satisfaction, just as with learning performance and engagement.

## 2.3 Learning Analytics Interventions

Interventions are considered crucial as a means to encouraging students who are at risk of failure, and also to help teachers make their lectures more effective (Na & Tasir, 2017). In general, interventions are used to change behaviors from undesirable to desirable in order to positively affect student learning. Interventions can also help provide just-in-time support to students (Wong & Li, 2020).

Many types of data have been used to determine the most appropriate intervention points, with students' online learning behaviors, study performance, demographics, and course selection information among the most preferred data used in the literature (Wong & Li, 2020). However, the existing studies have generally disregarded students video data. Clickstream interactions such as *play*, *pause*, *forward*, and *backward* can help provide a picture about students' futures according to the video lectures that they watch (Sinha et

al., 2014). They are also seen as useful in determining the correct timing for interventions so as to increase the effectiveness of video-based lectures. Timely and successful interventions will therefore help learners to benefit from the video lectures they view (Rienties, Boroowa, et al., 2016).

Most studies have focused on interventions such as the sending out of e-mails (Klein et al., 2019; Rienties, Cross, et al., 2016; Wong, 2017). However, such intervention types are not considered to be "just-in-time interventions," whilst watching video lectures naturally requires some form of immediate intervention. The existing studies have shown that the development and application of interventions for video-based learning can play an important role in the field of learning analytics, and thereby, further investigation is required in order to increase the effectiveness of video lectures through the development and employment of just-in-time interventions.

### 2.4 Interactive Videos

Many students do not complete their courses because of the requirement to watch straightforward linear video lectures (Chatti et al., 2016). Therefore, educators have started to focus on interactivity techniques in order to keep students more attentive and active during video-based lectures. Many studies have focused on the importance of interactive videos; however, it can be said that many MOOCs are still far from being interactive.

Suitable interactivity and intervention techniques can also be determined by tracking the watching behaviors of learners. Then, matching suitable interactivity techniques with different watching behaviors can help to increase the impact of a video lecture, as each interactivity technique is not necessarily appropriate for every type of watching behavior. Watching behaviors are considered useful input that may be used to make linear videos into interactive videos, and therefore existing linear videos in MOOCs could be transformed to become interactive. The most popular form applied involves the embedding of questions, with question types varied based on the context or purpose of the video. Multiple-choice, open-ended, and true-false questions are generally the most preferred (Wachtler, Khalil, et al., 2016). To date, few studies have investigated the effect of embedded questions in videos (e.g., Mar et al., 2017; Rice et al., 2019; Vural, 2013; Wachtler, Hubmann, et al., 2016).

Existing studies emphasized the positive effect of making interactive videos on students' learning (Cherrett et al., 2009). The study conducted by Zhang et al. (2006) showed that integrating interactive instructional videos into e-learning systems can accelerate students' learning and satisfaction. Petan et al. (2014) also concluded in their studies that interactive videos should be main resources of MOOCs.

## 2.5 Learning Analytics Framework

The number of studies that have aimed at developing a framework for learning analytics has increased recently. However, these frameworks have generally focused on the evaluation, implementation, or design architecture of learning analytics systems. That is, while some have evaluated systems according to predetermined dimensions or items, others have produced guidelines about how to implement learning analytics, as well as how to design the architecture of learning analytics systems.

An evaluation framework was developed by Scheffel (2017) to standardize the assessment of learning analytics tools, based on five dimensions (objectives, learning support, learning measures and output, data aspect, and organizational aspect), with each dimension including four items regarding consideration. Another framework was developed by Bakharia et al. (2016) to evaluate systems, and used a framework with five dimensions (temporal analytics, tool-specific analytics, cohort dynamics, comparative analytics, and contingency), with each dimension including specific metrics and visualizations in the evaluation of learning analytics systems.

In terms of implementation, two studies have attracted the attention of researchers. First, Greller and Drachsler (2012) introduced a framework of six dimensions (stakeholders, objective, data, instruments, external limitations, and internal limitations). The second framework, published by West et al. (2016), considered the implementation of learning analytics with a two-dimensional framework. The first dimension included the domains of institutional context, transitional institutional elements, learning analytics infrastructure, transitional retention elements, learning analytics for retention, intervention, and reflection, while the second dimension included a set of questions for each domain.

Two frameworks have focused on the design architecture of learning analytics systems. Fernández-Gallego et al. (2013) proposed a framework consisting of two components: 3D EVW platform and process mining system. The framework uses process mining algorithms based on system event logs to follow the learning flow of a course. Alternatively, the framework developed by Lu et al. (2017) proposed a system architecture showing a number of subsystems and services.

There have also been intervention-related studies in the literature that have focused on learning analytics, even though they use different terms such as model, engine, or approach rather than framework (Gong & Liu, 2019; Şahin & Yurdugül, 2019; Zhang et al., 2020).

Even though the frameworks proposed in the literature are considered valuable for implementing, evaluating, or designing learning analytics systems, they do not consider the video dimension. Since videos form the primary component of a MOOC, video

viewing behaviors may therefore play an important role in the development of a framework for video-based learning analytics. In addition, the existing studies have generally focused on frameworks at the macro level; that is, they have generally included institutional terms that are very hard for instructors to implement. Moreover, these frameworks include various constructs that have led to them being seen as complex. Differently from the existing literature, the current study focuses specifically on the development of a video-based learning analytics framework, from which instructors can improve their video lectures based on their application of the framework. The other advantage of the proposed framework is that it can be integrated into existing frameworks, since none as yet consider the video dimension. That is, it may be seen as a complementary framework for existing frameworks that currently disregard video watching behaviors and interactive videos.

## 2.6 Learning Variables in Online Courses

There have been a number of studies published with a focus on learning variables in online courses. Learner satisfaction, engagement, and academic performance are the primary variables investigated in such studies (Rajabalee & Santally, 2021), and they have searched for different factors affecting the learning variables (Muñoz-García & Villena-Martínez, 2021).

### 2.6.1 Satisfaction

Learner satisfaction is an important learning context when evaluating the efficiency of online courses in distance education, as well as having also been associated with learning performance (Kuo et al., 2014). In the current literature, it has been observed that the relationship between interaction types and learning satisfaction, as well as the relationship between variables of the Technology Acceptance Model (TAM) and satisfaction, have been especially investigated.

In a considerable number of studies, the effect of the primary types of interaction have been examined on learner satisfaction of online courses. In a study conducted by Gameel (2017), while learner-content interaction was found to be an important indicator of satisfaction, learner-learner and learner-instructor interactions were not found to have any direct relationship with learner satisfaction. In another study, Cho (2011) suggested learner-interface interaction as a means to increasing learner satisfaction, which is sometimes referred to as learner-technology interaction (Dzakira, 2003). Strachota (2003) reported that learner-content interaction was the most important indicator, while learner-instructor and learner-satisfaction interaction were of secondary importance as indicators of students' levels of satisfaction. A number of studies have emphasized the benefits of learner-instructor interaction on student satisfaction (Dennen et al., 2007; Keeler, 2006);

however, it has not been possible for instructors to provide this type of interaction in MOOCs due to the vast numbers of participants (Khalil & Ebner, 2015).

In addition to interaction-related studies, learner satisfaction for online courses has also been evaluated in studies with a focus on the (TAM) or similar variables of that model (Al-Azawei et al., 2017; Nagy, 2018). A significant numbers of studies have aimed to develop an extension of TAM to include learner satisfaction, with different names such as "perceived satisfaction" (Al-Azawei & Lundqvist, 2015). Studies have highlighted the importance of learner satisfaction in the acceptance of online courses.

In the wide variety of satisfaction-related studies published, their findings have shown satisfaction to be a variable requiring further investigation in order to increase the efficiency of online courses. Especially, if a study focuses on interactivity in online courses, it still stands out for researchers.

### 2.6.2 Engagement

The quality of learning in online courses can be assessed by observing learners' engagement (Shukor et al., 2014). Therefore, understanding engagement in online courses carries vital importance (Kahn et al., 2017). Accordingly, it is considered by researchers as one of the most preferred indicators for evaluating learning outcomes (Jung & Lee, 2018). However, the term "engagement" has been defined differently in the literature, and various studies have differentiated in their focus according to their general purpose. In order to clarify the meaning of engagement and to elaborate on engagement-related studies, Redmond et al. (2018) attempted to explain it in terms of a multidimensional perspective, and as such introduced five dimensions of engagement:

- Cognitive engagement
- Behavioral engagement
- Emotional engagement
- Social engagement
- Collaborative engagement

*Cognitive engagement* includes self-regulation techniques and metacognitive efforts (Wang & Eccles, 2012), and is seen as the most beneficial indicator for high learning performance (Dogan, 2015). *Behavioral engagement* relates to learner participation in learning activities (Buelow et al., 2018), whilst *Emotional engagement* concerns emotional reactions such as interest, boredom, happiness, sadness, and anxiety related to learning activities (Hu & Li, 2017). *Social engagement* means a student's desires and efforts to socialize (Daniels et al., 2016), and *Collaborative engagement* refers to a student's cooperation with their peers and instructor/s to increase learning (Ghaemi &

Lam, 2021), although there has been little research published about this dimension of engagement (Järvelä et al., 2016).

From these dimensions, cognitive, behavioral, and emotional are the most referred to dimensions of engagement in the literature (Park & Yun, 2018; Pellas & Kazanidis, 2015; Peng, 2017). Of course, different studies can utilize different dimensions depending on their research focus (Cho & Cho, 2014; Li & Baker, 2018). In addition, while some studies also use the term "interactivity" interchangeably with engagement in online courses (Northrup, 2001; Smith & Winking-diaz, 2004), many other studies decompose engagement from interactivity (Al-rahmi et al., 2015; Blasco-arcas et al., 2013).

The existing studies in the literature emphasize the need to explore students' engagement in order to make online courses work better (Czerkawski & Lyman, 2016), as well as revealing important findings about the factors that affect learner engagement.

### 2.6.3 Learning Performance

Learning performance or Academic performance refers to the learning outcome represented by a measurement score (Sullivan et al., 2017). It is also referred to as students' performance, grade, or achievement in the literature (Morris et al., 2005; Nakayama et al., 2007; Puzziferro, 2008). Academic performance shows to what extent learning goals have been attained in an online course. The number of studies that have focused on academic performance has been quite high when compared to other learning variables, since it is the most concrete and recognizable indicator of student success within online courses. The current studies have generally investigated relationships between different variables and academic performance in online courses in order to reveal which factors are the most affective on learners' academic performance (Coldwell et al., 2008; Gupta et al., 2013; Yeboah & Smith, 2016). All of these studies have shown that learning performance has been a major focus for researchers for many years, as well as still being current in the latest literature.

# CHAPTER 3

# RESEARCH METHOD

This chapter presents the research design of the study, and then explains each experiment in detail, with information provided on the participants, data collection experiments, procedures, and results introduced for each experiment.

## 3.1 Research Design

This study includes three experiments, with each focusing on different aims during the research. Figure 1 summarizes the three experiments, along with their aims, data collection instruments, and participants. It should be noted that the output of one experiment is then used as input to the next. In other words, the output of one experiment in used to shape the design of the subsequent one. Therefore, the final experiment (based on the one-group pretest-posttest design) is equipped to comprehensively answer the study's research questions.

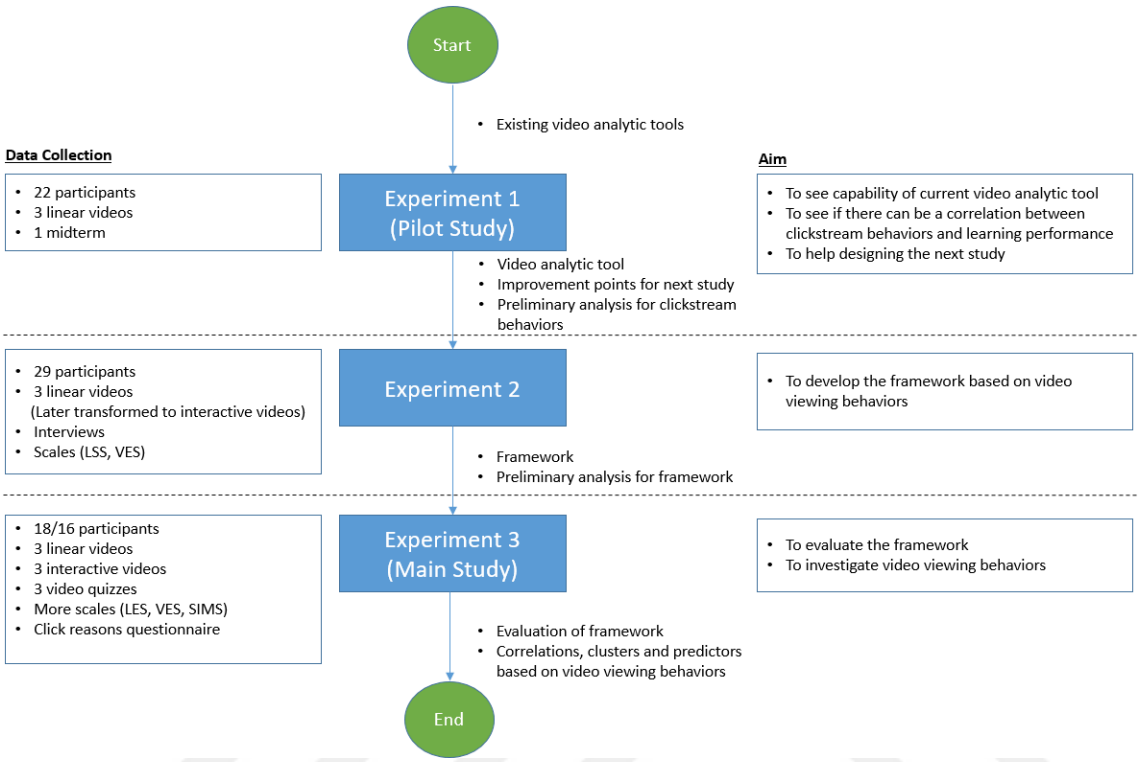



**Figure 1.** Research Design

*18/16 refers to 18 students' data having been used to evaluate the framework, whilst 16 students' data was used to investigate video clickstream behaviors with comprehensive learning analytics. See Section 4.2 for details.
*Each experiment uses different participants.

For all three experiments, convenience sampling (Fraenkel et al., 2012, p. 99) was used as the sampling method as the researcher had easy access to both the instructors and students from the department of Computer Education and Instructional Technology. Details regarding each experiments' participants are provided in Sections 3.2.1, 3.3.1, and 3.4.1.

## 3.2 Experiment 1 (Pilot Study)

The purpose of the pilot study was to observe the capability of the current video analytic tool, and to investigate whether or not any correlation exists between the students' watching behaviors and their learning performance. This first experiment was based on the one shot case study design defined by Fraenkel et al. (2012, p. 269).

### 3.2.1 Participants

The pilot study was performed within the Principles and Methods of Instruction course (CEIT 216) offered by the department of Computer Education and Instructional Technology (CEIT). There were 28 students enrolled to the course; hence the pilot study's participants were 28 second-grade CEIT students; However, only 22 had watched the video lectures.

### 3.2.2 Data Collection Instruments

Two types of data were collected during the pilot study: (1) Video clickstream behaviors and (2) Learning performance. The students' video clickstream behaviors were collected via a learning management system with an integrated video analytics tool. The students' midterm examination was conducted in a face-to-face classroom environment, and their learning performance was assessed based on their midterm scores.

#### 3.2.2.1 Learning Management System

Over a 3-week period, the participant students watched linear video lectures, and their clickstream interactions of *play*, *pause*, *backward*, and *forward* were collected via the learning management system. The following subsections explain the metrics and video lectures employed.

##### 3.2.2.1.1 Metrics

The participant students watched linear video lectures and their clickstream interactions were collected via SocialSkip, which is an open-source web application that collects video interaction data about watchers. It was integrated to odtuclass.metu.edu.tr, which is the primary learning management system of METU in order that TesterId could be matched from the student participants' username.

The list of metrics collected includes the students' viewing behaviors based on their clickstream interactions of *Play*, *Pause*, *Forward*, and *Backward.*

##### 3.2.2.1.2 Video Lectures

The participant students watched three different linear video lectures, with each focused on a famous educational phycologist. All of them (also including the video lectures in the other experiments) were developed by the same producer (Davidson Films). METU has a

license and videos are also located at METU Library. Table 3 details the content and duration of the video lectures:

**Table 3.** Video Lectures in Pilot Study

| Video Content | Duration (mm:ss) |
|---|---|
| **Piaget's Developmental Theory** | 27:06 |
| **Play: A Vygotskian Approach** | 26:13 |
| **Maria Montessori: Her Life and Legacy** | 35:10 |

### 3.2.2.2 Midterm

In order to measure the participant students' performance, a midterm examination was performed. The midterm consisted of 40 questions about the educational phycologists in the video lectures, with 35 multiple-choice and five true/false questions.

### 3.2.3 Procedure

In this pilot study, the students first watched three video lectures online via the university's learning management system, and then took a midterm exam in a face-to-face classroom setting. Figure 2 presents the steps that the students took during the pilot study:

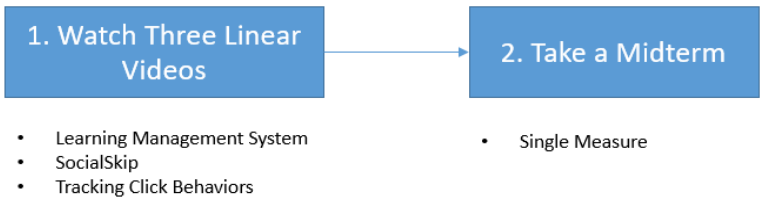



**Figure 2.** Pilot Study Steps

The participant students watched the three videos at their own pace as a part of the course. Then, 2 weeks after having watched the videos, the students took a midterm exam that accounted for 30% of their course assessment, with midterm scores were calculated out of 100 points for each student.

### 3.2.4 Results

The results of the pilot study are presented in two subsections since it targeted two different aims. One was related to a video analytics tool, whilst the other concerned correlations between clickstream behaviors and the students' learning performance.

#### 3.2.4.1 Results for Video Analytics Tool

Improvement points for a video analytics tool were determined from the pilot study's results. Since SocialSkip does not show *video dropout time*, *completion status*, and some basic interactions such as *volume change*, *mute*, and *full screen*, it could only be analyzed up to a certain degree.

The following metrics (interactions) were determined to be tracked by the system in the second experiment: *Open*, *Play*, *Paused*, *SeekForward*, *SeekBackward*, *VideoCompleted*, *ExitfromPlayer*, *ExitfromBrowser*, *ExitfromTimeout*, *Mute*, *Unmute*, *VolumeUp*, *VolumeDown*, *FullscreenOn*, *FullScreenOff*, and *Replay*.

These interactions should be saved in a *User Interaction Table* having the following attributes: *Interaction ID*, *Interaction Name*, *Parameters (Skip Time)*, *Video Current Time*, *Status (paused, playing)*, *User ID*, *Date*, *Video ID*, and *Order*.

#### 3.2.4.2 Results for Watching Behaviors

In the pilot study, the participant students each had a midterm exam score and video viewing behavior data regarding the three video lectures they had watched. Since one score exists for their academic performance, the total of the three video lectures was used in order to assess any correlation between the number of watching behaviors and the students' learning performance (midterm score).

A positive relationship was found to exist between *total number of interactions* (i.e., all interaction types) that the students performed and their midterm score (learning performance, $r = .59$, $p < .01$). In addition, Figure 3 illustrates the Pearson correlations between interaction numbers per interaction type and midterm score.

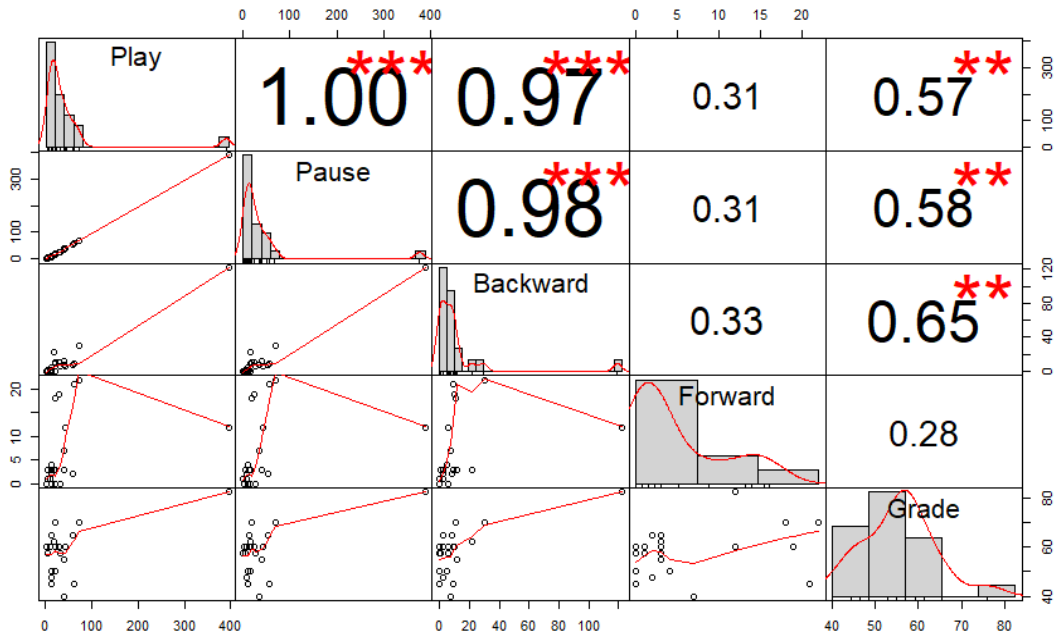



**Figure 3.** Pearson Correlation Matrix

The results also show that

- *Number of backwards* was found to be the most important indicator for predicting students' learning performance ($r$ = .65, $p$ < .01).
- A positive relationship was found to exist between *number of pauses* ($r$ = .58, $p$ < .01) and the students' learning performance.

### 3.2.4.3 Summary

The needs analysis showed that a video analytics tool or system having more metrics and attributes needed to be developed in order that students' watching behaviors could be tracked in a more appropriate and accurate way. Using SocialSkip, it is not currently possible to understand if a student stops watching a video or finishes it. Therefore, certain interactions to identify dropout were also determined. In addition, interactions with the video player such as *FullSreenOn* and *Mute* were also added in order to be sure as to whether or not students performed these interactions continuously or otherwise.

In terms of the students' video watching behaviors, the pilot study results revealed a positive relationship between the total number of interactions and learning performance. However, the pilot study was analyzed with only a single independent measure (one midterm score); therefore, it would be advisable to investigate the results with repeated measures.

### 3.3. Experiment 2

An experiment based on one-group pretest-posttest design was then performed in order to develop the framework based on the video viewing behaviors of students.

#### 3.3.1 Participants

The second experiment was conducted in a "Principles and Methods of Instruction" course that was offered to students studying in the CEIT department of a public university in Turkey. There were a total of 29 students enrolled to the course. The course was given during the fourth semester in the CEIT curriculum; therefore, the participant students were all second-grade undergraduate students, and none had been participants in the previous experiment (pilot study).

#### 3.3.2 Data Collection Instruments

##### 3.3.2.1 Video Lectures

The participant students viewed three different video lectures: (1) "Skinner: A Fresh Appraisal" (32:40 minutes), (2) "Piaget's Developmental Theory: An Overview" (27:06 minutes), and (3) "Play: A Vygotskian Approach" (26:13 minutes). Each video was about an educational psychologist, and all three videos were created by the same producer as in the other experiments, and with a common duration of around 30 minutes.

##### 3.3.2.2 Interviews

Interviews were conducted with both the participant students and a subject-matter expert.

The schedule for the students' interviews contained "why" questions, since the aim was to reveal the reasons behind the various click types recorded: for example, "Why did you click *pause*/*backward*/*forward* on time when the click intensity was high?"

The schedule for the subject-matter expert's interview included questions regarding revising the framework as the aim was to find ways to improve it: for example, "What do you think about the framework? How could we update the framework?"

##### 3.3.2.3 Scales

The "Learner Satisfaction Scale" (LSS) (Donkor, 2011) and the "Video Engagement Scale" (VES) (Visser et al., 2016) were used for the purposes of preliminary analysis in

this second experiment. The LSS consists of seven items rated on a 4-point, Likert-type scale, whilst the VES consists of 15 items rated on a 7-point, Likert-type scale. Both are validated and reliable scales. Cronbach's alpha value for LSS was found to be 0.88 in the study which was conducted for measuring its validity and reliability. In addition, Cronbach's alpha value for VES overall was found to be 0.93 and 0.94 in two different studies.

### 3.3.3 Procedure

The second experiment follows two distinct processes; the development of the framework, and the application of the framework; each of which are explained in the following subsections:

#### 3.3.3.1. Development of Framework

The framework to be developed was aimed to increase the effect of video-based learning through interventions. These interventions could be embedded based on the video watching behaviors of the students. Therefore, click types such as *pause*, *backward*, and *forward* are the location indicators from which interventions would be added. However, knowledge of click types alone, without understanding the reason for their usage, is not considered adequate to determine an intervention type. Therefore, if the reasons behind certain watching behaviors of students are better known, then suitable interventions could be developed and implemented.

Adding interventions to videos can be implemented along with interactive elements such as text, images, links, or questions. Since the content of interventions change according to the type of intervention, the actions (explain, exercise, motivate) contained in the interactive elements will also require changing. In order to match each of the categories (click types, reasons, interventions, actions, and interactive elements), an experimental study was conducted based on the following steps:

1. Students watched linear videos in a computer laboratory over a 3-week period.
2. Students' video clickstream behaviors analyzed.
3. Students interviewed regarding their video clickstream behaviors.
4. Students' interviews analyzed.
5. Framework developed based on interview results and the literature (Framework Version 1).
6. Interview conducted with subject-matter expert.
7. Draft framework revised (Framework Version 2).

*Step 1: Students watched linear videos in a computer laboratory over a 3-week period*

The participant students watched three linear videos (one video each week) in a computer laboratory, and their video clickstream behaviors were tracked using an online platform developed according to the pilot study's results. The platform includes the metrics mentioned in Section 3.2.4.1 Results for Video Analytics Tool. Whilst watching the videos, the students were permitted to take notes if they wanted. After having watched each video, a mini-quiz was applied to the students, and they each completed the Learner Satisfaction Scale and the Video Engagement Scale with regards to the video they had just watched.

The following are the videos that the students watched in this step:

1. "Skinner: A Fresh Appraisal" (32:40 minutes)
2. "Piaget's Developmental Theory: An Overview" (27:06 minutes)
3. "Play: A Vygotskian Approach" (26:13 minutes)

*Step 2: Students' video clickstream behaviors analyzed*

In this step, the video clickstream behaviors of the students were analyzed. The system tracked the students' interactions of *play*, *pause*, *forward*, *backward*, *volume up*, *volume down*, *mute*, *unmute*, *full screen on*, and *full screen off*. However, most students do not generally perform click interactions except for the most basic (*play*, *pause*, *forward*, and *backward*); therefore, this step focused especially on three types of interactions: *pause*, *forward*, and *backward*.

For each video, line graphs and box plots were created according to the recorded clickstream behaviors in order to find the peak points for each click type (see Appendix A and Appendix B).

Table 4 provides the numbers of interactions performed by the students who participated in the second experiment:

**Table 4.** Number of Interactions

| Video | Play | Pause | Backward | Forward | Total |
|---|---|---|---|---|---|
| **Skinner** | 701 | 670 | 153 | 483 | 2,186 |
| **Piaget** | 673 | 654 | 218 | 489 | 2,172 |
| **Vygotsky** | 1,224 | 1,200 | 99 | 723 | 3,357 |
| **Totals** | 2,598 | 2,524 | 470 | 1,695 | 7,715 |

*Step 3: Students interviewed regarding their video clickstream behaviors*

In order to establish the reasons behind the students' video clickstream behaviors, interviews were conducted with 17 of the participant students. Prior to the interviews, graphs were prepared for three types of clicks for each student: *pause*, *backward*, and *forward*. These graphs were then presented to the students in their interviews, and they were asked about the reasons behind the points having high click intensity. (Appendix C shows an example of the graphs shown to the interviewed students).

*Step 4: Interviews were analyzed*

The students' interviews were analyzed according to content analysis strategy in order to reveal the reasons for each click type. Content analysis strategy is explained in section 3.5. Table 5 presents the results of interview analysis:

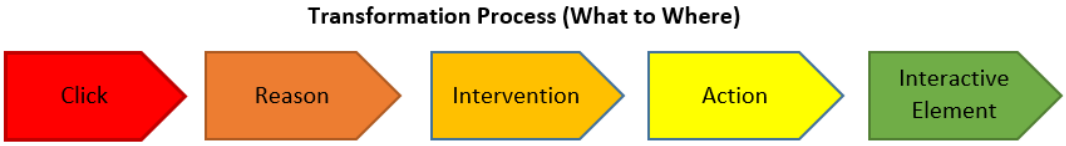
**Table 5.** Reasons of Watching Behaviors

| **Click** | **Reason** |
|---|---|
| **Pause** | Notetaking |
| **Backward** | Important points |
| | Language issues |
| | Not understanding |
| | Losing concentration |
| **Forward** | Unimportant points |
| | Finishing quickly |
| | Skimming |

*Step 5: Framework developed based on interview results and the literature (Version 1)*

Whilst developing the framework, the primary aim was to guide educators and researchers as to where certain elements should be placed in videos. In the literature, there are four important interactive elements (Bakla, 2017; Kleftodimos & Evangelidis, 2016b; Magdin et al., 2011): *Question*, *Text*, *Image*, and *Link.*

Accordingly, the literature presents three intervention types (Geller, 2005; Şahin, 2018): *Instructional, Supportive,* and *Motivational.*




| Click | Reason | Intervention | Action | Interactive Element |
|---|---|---|---|---|
| **Where** | **Why** | **Which** | **How** | **What** |
| Pause | Notetaking | Supportive | Explain | Text/Link |
| Backward | Not Understanding | | | |
| | Important Points | Instructional | Exercise | Question |
| | Language Proficiency | | | |
| | Speaker's Accent | | | |
| | Losing Concentration | Motivational | Motivate | Text/Image |
| Forward | Unimportant Points | | | |
| | Finishing Quickly | | | |
| | Skimming | Instructional | Exercise | Question |

**Figure 4.** Framework (First Version)

* This is not the finalized version of the framework for data collection (see Step 7).

The motivation behind the framework is to match suitable interactivity techniques with different video clickstream behaviors as a means to increasing the impact of video lectures, because each interactivity technique is not necessarily suited to every behavior type. As a result, the existing linear videos could be transformed into interactive videos.

*Step 6: Interview conducted with subject-matter expert*

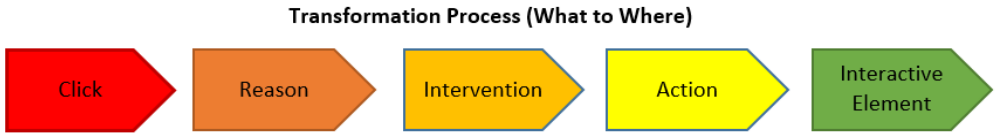
To improve the framework, an interview was conducted with a subject-matter expert with 9 years' experience with MOOCs courses and instructional technology. The subject matter expert has also a PhD on MOOCs from the department of Computer Education and Instructional Technology. The interview was made with a single subject matter expert since finding an available subject-matter expert on both subjects (MOOCs and instructional technology) was not easy and the selected subject-matter expert has very knowledgeable in this field. In other words, the most suitable person for the purpose was selected. Each element of the framework was discussed, and the process helped deliver improvements, especially in determining intervention types for each click reason.

*Step 7: Draft framework revised (Version 2)*

Following the subject-matter expert's interview, the framework was updated, and then click priority was added as a final measure in order to prevent confusion where different click types were performed at the same points of a video lecture. To add some clarity here, "click priority" means:

- *Forward* may support *backward*
- *Backward* may support *pause*

Therefore, where *pause* was followed by *backward* having been clicked at the same point in a video, "notetaking" was accepted as the reason for the clicks having been made by the user. Also, where *backward* was clicked followed by *forward* at the same point in a video, one of the reasons pertaining to *backward* is taken as the reason.




| Click | Reason | Intervention | Action | Interactive Element |
|---|---|---|---|---|
| **Where** | **Why** | **Which** | **How** | **What** |
| Pause | Notetaking | Supportive | Explain | Text/Link |
| Backward | Important Points | | | |
| | Language Issues | | | |
| | Not Understanding | Instructional | Exercise | Question |
| | Losing Concentration | Motivational | Motivate | Text/Image |
| Forward | Unimportant Points | | | |
| | Finishing Quickly | | | |
| | Skimming | No Need for Intervention | - | - |

High — Click Priority — Low

**Figure 5.** Framework (Final Version)

### 3.3.3.2. Application of Framework

In order to apply the framework, interactive videos were first developed based on the framework by using h5p, an open source community-driven project. During this period, interventions were added according to the framework in order to make the videos "interactive." Example interventions added to the video lectures are presented in Appendix G. Then, the same students who participated in the development phase watched the interactive videos 1 month later.

Table 6 presents the numbers of interventions added to each video lecture, which was determined according to the peak points of three different click types (*pause*, *backward*, and *forward*).

**Table 6.** Number of Interventions

| Video | Number of Interventions |
|---|---|
| Skinner: A Fresh Appraisal | 11 |
| Piaget's Developmental Theory: an Overview | 16 |
| Play: A Vygotskian Approach | 14 |
| **Total** | 41 |

The participant students watched the online interactive videos developed, and were then asked to complete the following scales:

1. Intervention Surveys (Timing, Content)
2. Satisfaction Scale
3. Engagement Scale

### 3.3.4 Results (Preliminary Evaluation of Framework)

The framework was evaluated in terms of satisfaction and engagement. Therefore, the Video Engagement Scale (Visser et al., 2016) and the Learner Satisfaction Scale (Donkor, 2011) were given to each student to complete. The scale results from the interactive videos were then compared to those from the linear videos.

Whilst watching the interactive videos, the students analyzed each intervention in terms of its timing and content, and then evaluated the effect of the framework in terms of their engagement and satisfaction.

#### 3.3.4.1 Analysis of interventions

Each intervention is evaluated in terms of timing and content. *Timing* refers to showing intervention in correct time, whereas *Content* refers to the suitability of the intervention to video content at that moment. Wachtler, Khalil, et al. (2016) emphasized the importance of evaluating timing and content of interventions in interactive videos. The following subsections explain the results in terms of both timing and content for each intervention in each video. There were two items in the survey (one for timing, one for content) for

each intervention. The instruction given was “Please assess each intervention in terms of timing and content,” using a rating scale from 1 (*very bad*) to 5 (*very good*).

#### 3.3.4.1.1. Skinner

In total, 11 interventions were added according to the peak points on the graphs for *pause*, *backward*, and *forward* related to the video lecture about Skinner. Of those 11, seven interventions were supportive, three were motivational, and one intervention was instructional. Table 7 presents the results for each intervention in the Skinner video lecture:

**Table 7.** Survey Results for Each Intervention (Video Lecture 1)

| Intervention Order | Type | Timing | Content | Mean |
|---|---|---|---|---|
| 1 | Supportive | 3.96 | 4.21 | 4.09 |
| 2 | Motivational | 4.04 | 3.68 | 3.86 |
| 3 | Supportive | 4.61 | 4.71 | 4.66 |
| 4 | Instructional | 4.07 | 4.39 | 4.23 |
| 5 | Supportive | 4.25 | 4.32 | 4.29 |
| 6 | Supportive | 4.50 | 4.50 | 4.50 |
| 7 | Motivational | 3.68 | 3.43 | 3.55 |
| 8 | Supportive | 4.46 | 4.50 | 4.48 |
| 9 | Supportive | 4.61 | 4.75 | 4.68 |
| 10 | Supportive | 4.54 | 4.71 | 4.63 |
| 11 | Motivational | 3.86 | 3.57 | 3.71 |
| **Mean** | | 4.23 | 4.25 | 4.24 |

Table 8 presents the results for each type of intervention in the Skinner video lecture according to the mean scores for each intervention type:

**Table 8.** Results for Each Type of Intervention (Video Lecture 1)

| Type | Timing | Content | Mean |
|---|---|---|---|
| Motivational | 3.86 | 3.56 | 3.71 |
| Instructional | 4.07 | 4.39 | 4.23 |
| Supportive | 4.42 | 4.53 | 4.47 |

As can be seen from Table 8, supportive interventions achieved the highest scores, whilst the motivational interventions received the lowest among the intervention types. Supportive (Mean = 4.47) and instructional interventions (Mean = 4.23) were evaluated as "good" or "very good," whereas motivational interventions (Mean = 3.71) were "neither good nor bad" or "good." However, all were above the mid-point (Mean = 3.00) and near to or above the mean point representing good (Mean = 4.00). In addition, the mean scores for both timing (Mean = 4.23) and content (Mean = 4.25) for all interventions are considered to be high (see also Table 7).

#### 3.3.4.1.2. Piaget

By examining the peak points, there were 16 interventions added to the video lecture related to Piaget. Of those 16 interventions, 11 were supportive interventions, three were motivational, and two of them were instructional interventions. The scores for timing and content can be seen as shown in Table 9:

**Table 9.** Results for Each Intervention (Video Lecture 2)

| Intervention Order | Type | Timing | Content | Mean |
|---|---|---|---|---|
| 1 | Motivation | 4.04 | 3.92 | 3.98 |
| 2 | Supportive | 4.19 | 4.15 | 4.17 |
| 3 | Supportive | 4.31 | 4.50 | 4.40 |
| 4 | Supportive | 4.35 | 4.46 | 4.40 |
| 5 | Supportive | 4.27 | 4.31 | 4.29 |
| 6 | Instructional | 4.46 | 4.50 | 4.48 |
| 7 | Supportive | 4.31 | 4.19 | 4.25 |
| 8 | Supportive | 4.31 | 4.38 | 4.35 |
| 9 | Supportive | 4.42 | 4.42 | 4.42 |
| 10 | Instructional | 4.35 | 4.46 | 4.40 |
| 11 | Motivational | 3.69 | 3.54 | 3.62 |
| 12 | Motivational | 3.69 | 3.65 | 3.67 |
| 13 | Supportive | 4.35 | 4.19 | 4.27 |
| 14 | Supportive | 4.31 | 4.27 | 4.29 |
| 15 | Supportive | 4.35 | 4.23 | 4.29 |
| 16 | Supportive | 4.50 | 4.50 | 4.50 |
| **Mean** | | **4.24** | **4.23** | **4.24** |

Table 10 shows the results for each type of intervention in the Piaget video lecture according to the mean scores for each intervention type:

**Table 10.** Results for Each Type of Intervention (Video Lecture 2)

| Type | Timing | Content | Mean |
|---|---|---|---|
| Motivational | 3.81 | 3.71 | 3.76 |
| Instructional | 4.40 | 4.48 | 4.44 |
| Supportive | 4.33 | 4.32 | 4.33 |

As can be seen from Table 10, the instructional interventions received the highest scores, whilst the motivational interventions were the lowest among the intervention types. The instructional (Mean = 4.44) and supportive interventions (Mean = 4.33) were evaluated as "good" or "very good," whereas the motivational interventions (Mean = 3.76) were evaluated as "neither good nor bad" or "good." However, all were above the mid-point (Mean = 3.00) and near to or above the mean point representing good (Mean = 4.00). In addition, the mean scores for both timing (Mean = 4.24) and content (Mean = 4.23) for all interventions were considered as being high (see also Table 9).

### 3.3.4.1.3. Vygotsky

The video lecture related to Vygotsky included 14 interventions (eight supportive, three instructional, and three motivational) according to an analysis of the peak points. The timing and content scores for each intervention are presented as shown in Table 11:

**Table 11.** Results for Each Intervention (Video Lecture 3)

| Intervention Order | Type | Timing | Content | Mean |
|---|---|---|---|---|
| 1 | Motivational | 4.05 | 3.64 | 3.84 |
| 2 | Supportive | 4.41 | 4.41 | 4.41 |
| 3 | Supportive | 4.41 | 4.50 | 4.45 |
| 4 | Supportive | 4.45 | 4.36 | 4.41 |
| 5 | Supportive | 4.45 | 4.50 | 4.48 |
| 6 | Instructional | 4.41 | 4.36 | 4.39 |
| 7 | Instructional | 4.27 | 4.41 | 4.34 |
| 8 | Supportive | 4.23 | 4.32 | 4.27 |
| 9 | Instructional | 4.36 | 4.36 | 4.36 |
| 10 | Supportive | 4.55 | 4.64 | 4.59 |
| 11 | Supportive | 4.41 | 4.50 | 4.45 |
| 12 | Supportive | 4.41 | 4.50 | 4.45 |
| 13 | Motivational | 3.82 | 3.91 | 3.86 |
| 14 | Motivational | 3.73 | 3.59 | 3.66 |
| **Mean** | | 4.28 | 4.29 | 4.28 |

Table 12 shows the results for each type of intervention in the Vygotsky video lecture according to the mean scores for each intervention type:

**Table 12.** Results for Each Type of Intervention (Video Lecture 3)

| Type | Timing | Content | Mean |
|---|---|---|---|
| Motivational | 3.86 | 3.71 | 3.79 |
| Instructional | 4.35 | 4.38 | 4.36 |
| Supportive | 4.41 | 4.47 | 4.44 |

As can be seen from Table 12, the supportive interventions received the highest scores, whilst the motivational interventions were the lowest among the intervention types. The supportive (Mean = 4.44) and instructional interventions (Mean =4.36) were evaluated as "good" or "very good," whereas the motivational interventions (Mean =3.79) were evaluated as "neither good nor bad" or "good." However, all were found to be above the mid-point (Mean =3.00) and near to or above the mean point representing good (Mean =4.00). In addition, the mean scores for both timing (Mean =4.28) and content (Mean =4.29) for all interventions were considered to be high (see also Table 11).

#### 3.3.4.2. Evaluation of Framework

In addition to the analysis of the interventions, the proposed framework was also subjected to preliminary analysis in terms of the students' satisfaction and engagement. Paired *t*-tests were performed in order to compare the mean values prior to and following the interventions, if certain assumptions were met. Therefore, the assumptions for paired *t*-test were also checked, which were that: (1) Subjects must be independent, (2) Pairs should be from the same participants, and (3) Differences among pretests and posttests must be normally distributed. The first two assumptions were actually the reasons for choosing paired samples *t*-test for comparison. Therefore, only Shapiro-Wilk results are reported for the third assumption in order to show that the assumptions were met for the paired samples *t*-test in the analysis. Wilcoxon signed-rank test was used if the assumptions were not met for a paired samples *t*-test.

For each video lecture, the results are reported in the following subsections:

##### 3.3.4.2.1. Skinner

To compare the mean satisfaction scores obtained for Video Lecture 1 (subject: Skinner), assumptions were controlled, and Shapiro-Wilk test showed that the measured differences were normally distributed ($p > 0.05$).

Table 13 shows the paired *t*-test results after comparing the mean satisfaction scores measured before and after the interventions added to Video Lecture 1:

**Table 13.** Comparison of Mean Scores for Satisfaction in Experiment 2 (Video Lecture 1)

| Video Type | *n* | $\bar{X}$ | *SD* | *df* | *t* | *p* |
|---|---|---|---|---|---|---|
| **Linear** | 25 | 2.89 | 0.65 | 24 | .557 | .583 |
| **Interactive** | 25 | 2.83 | 0.69 | | | |

The paired samples *t*-test showed that the interventions did not elicit a statistically significant change in the students' satisfaction for Video Lecture 1.

For comparing the mean engagement scores obtained for Video Lecture 1, assumptions were controlled and the Shapiro-Wilk test results showed that the measured differences were normally distributed ($p > .05$).

Table 14 presents the *t*-test results after comparing the mean engagement scores measured before and after interventions in Video Lecture 1:

**Table 14.** Comparison of Mean Scores for Engagement in Experiment 2 (Video Lecture 1)

| Video Type | *n* | $\bar{X}$ | *SD* | *df* | *t* | *p* |
|---|---|---|---|---|---|---|
| **Linear** | 25 | 4.29 | 1.22 | 24 | .282 | .780 |
| **Interactive** | 25 | 4.22 | 1.61 | | | |

The paired samples *t*-test showed that the interventions did not elicit a statistically significant change in the students' engagement for Video Lecture 1.

### 3.3.4.2.2. Piaget

The Shapiro-Wilk test results for student satisfaction for Video Lecture 2 (subject: Piaget) showed that the measured differences were normally distributed ($p > 0.05$).

Table 15 presents the *t*-test results after comparing the mean student satisfaction scores measured before and after the interventions for Video Lecture 2:

**Table 15.** Comparison of Mean Scores for Satisfaction in Experiment 2 (Video Lecture 2)

| Video Type | *n* | $\bar{X}$ | *SD* | *df* | *t* | *p* |
|---|---|---|---|---|---|---|
| **Linear** | 25 | 2.91 | 0.59 | 24 | .927 | .363 |
| **Interactive** | 25 | 2.83 | 0.73 | | | |

The paired samples *t*-test showed that the interventions did not elicit a statistically significant change in the students' satisfaction for Video Lecture 2.

The Shapiro-Wilk test results for engagement in Video Lecture 2 showed that the measured differences were not found to be normally distributed ($p < .05$). Therefore, the Wilcoxon signed-rank test was used for the purposes of comparison.

Table 16 shows mean engagement scores for Video Lecture 2:

**Table 16.** Comparison of Mean Scores for Engagement in Experiment 2 (Video Lecture 2)

| Video Type | *n* | $\bar{X}$ | *SD* | *Z* | *p* |
|---|---|---|---|---|---|
| **Linear** | 25 | 4.54 | 1.19 | 131 | .587 |
| **Interactive** | 25 | 4.39 | 1.70 | | |

The Wilcoxon signed-rank test showed that the interventions did not elicit a statistically significant change in the students' engagement for Video Lecture 2.

### 3.3.4.2.3. Vygotsky

The Shapiro-Wilk test results for student satisfaction in Video Lecture 3 (subject: Vygotsky) showed that the measured differences were normally distributed ($p > .05$).

Table 17 shows the paired *t*-test results after comparing the mean satisfaction scores measured before and after interventions in Video Lecture 3:

**Table 17.** Comparison of Mean Scores for Satisfaction in Experiment 2 (Video Lecture 3)

| Video Type | *n* | $\bar{X}$ | *SD* | *df* | *t* | *p* |
|---|---|---|---|---|---|---|
| **Linear** | 22 | 2.79 | 0.72 | 21 | .563 | .579 |
| **Interactive** | 22 | 2.64 | 0.78 | | | |

The paired samples *t*-test showed that the interventions did not elicit a statistically significant change in the students' satisfaction for Video Lecture 3.

The Shapiro-Wilk test results for engagement in Video Lecture 3 showed that the measured differences were normally distributed ($p > .05$).

Table 18 presents the *t*-test results after comparing the mean engagement scores measured before and after the interventions in Video Lecture 3:

**Table 18.** Comparison of Mean Scores for Engagement in Experiment 2 (Video Lecture 3)

| Video Type | *n* | $\bar{X}$ | *SD* | *df* | *t* | *p* |
|---|---|---|---|---|---|---|
| **Linear** | 22 | 3.96 | 1.51 | 21 | -.855 | .402 |
| **Interactive** | 22 | 4.20 | 1.80 | | | |

The paired samples *t*-test showed that the interventions did not elicit a statistically significant change in the students' engagement in Video Lecture 3.

#### 3.3.4.3. Summary

The results show that the timing and content for interventions were found to be high. However, the interventions did not elicit a statistically significant change in either the students' satisfaction or their engagement in the three video lectures. This may be because the framework does not work satisfactorily for all students. It may, however, prove useful for those who require interventions; hence, it would be prudent to investigate the effect of additional variables, especially motivation.

## 3.4 Experiment 3 (Main Study)

The research method employed for the main study was quasi-experimental. The data was collected from one group and dependent variables were then measured with both a pretest and posttest. Therefore, the study was conducted according to the one-group pretest-posttest design, which is a type of quasi-experimental study (Field & Hole, 2003). The effect of the intervention framework developed in Experiment 1 (pilot study) was investigated with interactive videos during this experiment (main study).

### 3.4.1 Participants

The participants of the main study were students enrolled to the EDS212 (Instructional Principles and Methods) course during the 2019/2020 Fall semester. There were a total of 22 registered students, and none had previously taken part in either the first or second experiments. However, two of the students did not continue with the course. In total, 18 students watched both the linear and interactive videos.

### 3.4.2 Data Collection Instruments

#### 3.4.2.1 Videos

<u>Linear Videos</u>

There were three linear videos provided during the course. Whilst watching each video, the students' watching behaviors were tracked via the online platform. These behaviors were based on the students' click interactions of *play*, *pause*, *forward*, and *backward*.

Table 19 details the three linear videos that the students watched:

**Table 19.** Linear Video Lectures in Main Study

| Video Content | Duration (mm:ss) |
|---|---|
| **Maria Montessori: Her Life and Legacy** | 35:10 |
| **Erik H. Erikson: A Life's Work** | 30:42 |
| **John Dewey: His Life and Work** | 29:55 |

Interactive Videos

The interactive videos were developed based on the framework, and included supportive, motivational, and instructional interventions. These interventions were each created according to the framework using h5p.

There were three interactive videos developed in the study, as shown in Table 20:

**Table 20.** Interactive Video Lectures in Main Study

| Video Content | Duration (mm:ss) |
|---|---|
| **Skinner: A Fresh Appraisal** | 32:40 |
| **Piaget's Developmental Theory: an Overview** | 27:06 |
| **Play: A Vygotskian Approach** | 26:13 |

### 3.4.2.2 Scales

In addition to applying the satisfaction (LSS) and engagement scales (VES) used in exploratory experiment, the "Situational Motivation Scale" (SIMS) (Guay et al., 2000) was also added to this experiment (see Appendix E). The intrinsic motivation subscale of SIMS was used in order to investigate the effect of the framework on the students' satisfaction, engagement, and learning performance by taking their motivation into account. The subscale contains a total of 4 items rated on a 7-point, Likert-type scale. Reliability and validity of the subscale were also ensured with different studies conducted by Guay, Vallerand, and Blanchard (2000). During this study, whenever a scale was filled, an overall scale score was calculated and used as a scale score in the data analysis since each scale was reliable and validated.

#### 3.4.2.3 Questionnaire

The “Click Reasons Questionnaire” was used to measure the impact of each intervention type, and to investigate changes regarding the students’ intervention needs (see Appendix F). The questionnaire was created based on the reasons given for interventions in the framework, and consisted of seven items rated on a 5-point, Likert-type scale.

#### 3.4.2.4 Video Quizzes

Six video quizzes were presented to the participant students, with one quiz administered following each video lecture. Each quiz consisted of eight or nine content-related multiple-choice or true/false questions. The students’ learning performance was assessed based on the students’ scores obtained from the video quizzes.

### 3.4.3 Procedure

The experiment was conducted over a 6-week duration. The focus of the first 3 weeks was the linear videos, whilst the second 3 weeks were for the interactive videos. The following six steps were applied in conducting this experiment:

During first three weeks, each week respectively:

1. Students watched a linear video at home.
2. Students completed the scales and the questionnaire at home (pretest).
3. Students received a quiz in the classroom (pretest).

During last three weeks, each week respectively:

4. Students watched an interactive video at home.
5. Students completed the scales and the questionnaire at home (posttest).
6. Students received a quiz in the classroom (posttest).

The scales and the questionnaire applied in Step 2 and the quizzes in Step 3 were used as pretests, whereas the scales and the questionnaire applied in Step 5 and the quizzes in Step 6 were used as posttests.

In total, 97 observations (44 from linear videos, 53 from interactive videos) were collected from the 18 participant students (some students did not watch all of the video lectures).

### 3.5 Data Analysis

During this research, different data analysis techniques were performed. Whilst basic statistical techniques were generally used in the first two experiments, various statistical techniques, including some that were more comprehensive, were performed in the third experiment (as the main study) which aimed to answer all six research questions.

In the first experiment (pilot study), Pearson correlation was used as a statistical technique since it is the most suited method to measuring bivariate linear relationships among continuous variables (Field, 2018). In the second experiment, during the framework's development process, content analysis strategy was employed to analyze the collected interview data since it aims to extract meaningful categories or themes from text or speech (Zhang & Wildemuth, 2005). In addition, descriptive statistics such as mean were used to individually evaluate the interventions in each video lecture. Moreover, *t*-test and Wilcoxon signed-rank test were selected as parametric and nonparametric tests respectively for the comparison of mean values from the pretest and posttest results as the preliminary analysis of the framework.

In the third experiment (main study), various data analysis techniques were employed in order to answer the six research questions of the study, and to address a total of 27 hypotheses. Since the pretests and posttests were obtained from 18 students (one group) and the aim was to compare the mean pretest-posttest scores of the learning variables (i.e., learning performance, satisfaction, and engagement), paired samples *t*-test was used to the investigation of the hypotheses in the first research question (RQ1). In addition, average scores were obtained from the pretests and posttests and these scores were compared using *t*-test or Wilcoxon signed-rank test in line with the pretest-posttest research design, as identical videos (three linear and three interactive) were used as the data collection instruments.

Moreover, it was seen that a linear relationship existed between motivation and satisfaction, engagement, and also learning performance. Therefore, the effect of motivation on the learning variables was investigated according to the hypotheses in RQ2. For this purpose, motivation was used as a covariate and ANCOVA was selected as the most suitable statistical technique where assumptions were met. Quade's ANCOVA was selected for those not meeting assumptions for parametric ANCOVA.

RQ3 examined the reasons behind the observed behaviors in the framework, and paired samples *t*-test and Wilcoxon signed-rank test were used to compare the mean scores of the single group's pretests and posttests. First, the assumptions for a paired sample *t*-test were investigated, and the test were performed if they were met. Else, the Wilcoxon signed-rank test was used since it is a common non-parametric version of *t*-test used to compare

paired samples. In the first research question, average scores were compared for investigating reasons.

Furthermore, repeated measure correlation analysis (rmcorr) was performed in addressing RQ4 as repeated measures were collected from the participants, and also that the participants' video clickstream behaviors may change on any given date or from one video to another.

For clustering, two-step cluster analysis and hierarchical clustering were employed in the study. First, two-step cluster analysis was preferred since it merges the principles of traditional clustering and returns an overall goodness of fit measure (Polier, 2019). In the case of an overall goodness of fit and cluster graphs not being found satisfactory with the two-step cluster analysis, hierarchical clustering was employed as the clustering method. Moreover, a student performs more than one observation (data point) and each observation was clustered. Clustering is a method that provides categorical data, and the observations of each student might not be treated as independent and residuals may be non-normal based on categorical data. Therefore, Robust linear mixed model was used to investigate the hypotheses in RQ5 which aimed to examine the group (cluster) effect.

In order to investigate the hypotheses in the last research question (RQ6), XGBoost and Robust LMM were selected as prediction algorithms. There are significant reasons for having preferring XGBoost and Robust LMM in the current study. XGBoost was the preferred algorithm since it is known to be scalable and effective when using gradient boosting trees, and also that it is considered to work well with small-sized datasets (Chen et al., 2015). In addition, Robust LMM was also preferred for the purposes of prediction since it does not make any assumptions with the exception that the model parameters can be estimated, and also that it is suited to repeated measures (Koller, 2016). Small samples especially have the potential to include outliers or non-Gaussian distributions (Demidenko, 2013); hence, Robust LMM is able to be used for non-normal distributed residuals when assumptions are not met for LMM (Schielzeth et al., 2020). In addition, three different evaluation schemes (Random Split, One Person Leave Out, and One Video Leave Out) were applied in evaluating the model. In random split evaluation scheme, data in a split was used for testing the model whilst data from all other splits was used for the purposes of training and validating the model. This was done for each split until all splits were used for testing. By doing so, both the test data was made unseen and all dataset was used for testing purposes. In one person leave out, a person's data was used instead of a split this time. Similarly, this was done for each person until data of all people was used for testing. In one video leave out, data for each video was used for testing the model while data from all other people was applied for training and validating the model. This was also performed for each video until data of all videos was used for testing.

All of the analysis methods employed to answer the study's research questions are explained in detail during Chapter 4.

# CHAPTER 4

# RESULTS

The study's findings are presented in two sections: Framework and Comprehensive Learning Analytics, since the study focuses on both framework and video clickstream behaviors with comprehensive learning analytics. As mentioned in Section 1.4, the study is formulated around six research questions, with three related to the framework and three on watching behaviors. The findings are reported against each research question in the following sections. During this section, the term "interactive videos" was shortly used for the term "interactive videos created according to the framework".

## 4.1 Results for Framework

The effect of the framework on the participant students' satisfaction, engagement, learning performance, and click reasons are presented in this section. The differences in these variables were examined both by not controlling for student motivation and also by controlling for motivation in accordance with the first and second research questions, respectively. For the third question, the impact of the proposed framework on the students' click reasons was inspected.

### 4.1.1 Differences with Motivation Not Controlled

*RQ1:* Are there any significant differences between linear and interactive videos in terms of students' satisfaction, engagement and learning performance?

The first research question includes three hypotheses (H1, H2, H3) in order to yield results regarding the students' satisfaction, engagement and learning performance separately.

#### 4.1.1.1 Satisfaction

*Hypothesis 1:* There is a significant difference between linear and interactive videos in terms of the students' satisfaction.

*Null hypothesis:* There is no significant difference between linear and interactive videos in terms of the students' satisfaction.

The Shapiro-Wilk test results for the students' satisfaction showed that the measured differences were normally distributed ($p$ >.05). Therefore, the students' satisfaction scores were compared with a paired samples *t*-test. Table 21 presents the comparison results for satisfaction:

**Table 21.** Comparison of Mean Scores for Satisfaction

| Video Type | *n (student)* | $\bar{X}$ | *SD* | *df* | *t* | *p* |
|---|---|---|---|---|---|---|
| **Linear** | 18 | 2.54 | 0.69 | 17 | 0.281 | .782 |
| **Interactive** | 18 | 2.50 | 0.50 | | | |

The results show that no significant difference was found between the linear ($M$ = 2.54, $S$ = 0.69) and interactive videos ($M$ = 2.50, $SD$ = 0.50) in terms of the students' satisfaction, $t(17) = 0.281$, $p > .05$.

### 4.1.1.2 Engagement

*Hypothesis 2:* There is a significant difference between linear and interactive videos in terms of students' engagement.

*Null hypothesis:* There is no significant difference between linear and interactive videos in terms of students' engagement.

The Shapiro-Wilk test results for engagement showed that the measured differences were normally distributed ($p$ > .05). The students' engagement scores were compared with a paired samples *t*-test. Table 22 illustrates the comparison results for engagement:

**Table 22.** Comparison of Mean Scores for Engagement

| Video Type | *n (student)* | $\bar{X}$ | *SD* | *df* | *t* | *p* |
|---|---|---|---|---|---|---|
| **Linear** | 18 | 3.40 | 0.99 | 17 | 1.985 | .070 |
| **Interactive** | 18 | 3.00 | 0.50 | | | |

The results show that no significant difference was found between the linear ($M$ = 3.40, $S$ = 0.99) and interactive videos ($M$ = 3.00, $SD$ = 0.50) in terms of students' engagement, $t(17) = 1.94$, $p > .05$.

#### 4.1.1.3 Learning Performance

*Hypothesis 3:* There is a significant difference between linear and interactive videos in terms of students' learning performance.

*Null hypothesis:* There is no significant difference between linear and interactive videos in terms of students' learning performance.

The Shapiro-Wilk test results for learning performance showed that the measured differences were normally distributed ($p > .05$). The learning performance scores were compared with a paired samples *t*-test. Table 23 presents the comparison results for learning performance:

**Table 23.** Comparison of Mean Scores for Learning Performance

| Video Type | *n (student)* | $\bar{X}$ | *SD* | *df* | *t* | *p* |
|---|---|---|---|---|---|---|
| **Linear** | 18 | 84.56 | 13.20 | 17 | 0.066 | .948 |
| **Interactive** | 18 | 84.38 | 7.50 | | | |

The results show that no significant difference was found between the linear ($M = 84.56$, $SD = 13.20$) and interactive videos ($M = 84.38$, $SD = 7.50$) in terms of students' learning performance, $t(17) = 0.66$, $p > .05$.

### 4.1.2 Differences with Motivation Controlled

*RQ2:* Are there any significant differences between linear and interactive videos in terms of students' satisfaction, engagement, and learning performance when motivation is controlled?

The second research question also includes three hypotheses (H4, H5, H6) in order to yield results for satisfaction, engagement, and learning performance separately when motivation is controlled.

Analysis of covariance (ANCOVA) was selected to compare linear and interactive videos by using motivation as a covariate since motivation is known to be positively related to e-learning (Harandi, 2015) as well as it was observed that motivation was positively related to the learning variables in this study. The relationship between motivation and each learning variable was also reported in following subsections.

ANCOVA has seven assumptions including generic assumptions and ANCOVA specific assumptions (Li & Chen, 2019, p. 201): (1) Independence of observations, (2) Absence of outliers, (3) Normality of residuals, (4) Homogeneity of variance, (5) Homogeneity of

regression slopes, (6) Linearity between covariate(s) and dependent variable, (7) No strong correlation among covariates. Assumptions were checked to be sure of usage of ANCOVA. Since there is one covariate (motivation), there is no need to check last assumption. In addition, assumptions (3), (4), (5), and (6) were reported in each subsection because observations were independent and outliers were omitted from the analysis if existed. In cases where parametric ANCOVA assumptions were not met, it was decided to use Quade (1967)'s ANCOVA.

#### 4.1.2.1 Satisfaction

*Hypothesis 4:* There is a significant difference between linear and interactive videos in terms of students' satisfaction scores when motivation is controlled.

*Null hypothesis:* There is no significant difference between linear and interactive videos in terms of students' satisfaction scores when motivation is controlled.

Differently from the first research question, 97 observations (44 from linear videos, 53 from interactive) from 18 students were individually used to compare the linear and interactive videos to control the effect of the students' motivation on each observation.

Firstly, assumptions of ANCOVA for satisfaction were controlled and Kolmogorov-Smirnov test ($n_{observation}>50$) was checked for normality of residuals assumption

Here are the results for the assumptions in terms of satisfaction:

- Kolmogorov-Smirnov test showed normality of residuals was met ($p>.05$).
- Levene's test indicated variances are equal ($p>.05$)
- There was a positive correlation between motivation and satisfaction ($r=.82$, $p=<.01$)
- Test for homogeneity of regression slopes showed the assumption was not violated ($p>.05$)

So, assumptions for ANCOVA were met for satisfaction. As a result, ANCOVA was used where motivation was selected as a covariate, and user identifier was selected as a random factor since users have more than one observation. Then, the main effect of video type (linear or interactive) based on estimated marginal means was compared in terms of satisfaction. Table 24 illustrates the comparison results for satisfaction when motivation was controlled:

**Table 24.** Comparison of Mean Scores for Satisfaction (when motivation was controlled)

| Video Type | *n* (Observation) | Mean | *SD* | Mean (adjusted) | *SE* | *F* | eta$^2$ |
|---|---|---|---|---|---|---|---|
| **Linear** | 44 | 2.54 | 0.79 | 2.43 | 0.06 | 4.440* | 0.069 |
| **Interactive** | 53 | 2.52 | 0.66 | 2.59 | 0.05 | | |

*$p$ = .039 < .05

The results show that a significant difference was found between the linear ($M$ = 2.43, $SD$ = 0.06) and interactive videos ($M$ = 2.59, $SD$ = 0.06) in terms of students' satisfaction scores when motivation was controlled, $p$ = .039 < .05.

### 4.1.2.2 Engagement

*Hypothesis 5:* There is a significant difference between linear and interactive videos in terms of students' engagement scores when motivation is controlled.

*Null hypothesis:* There is no significant difference between linear and interactive videos in terms of students' engagement scores when motivation is controlled.

The assumptions of ANCOVA for engagement were firstly controlled. Here are the results for the assumptions in terms of engagement:

- Kolmogorov-Smirnov test showed normality of residuals was met ($p$>.05).
- Levene's test indicated variances are equal ($p$>.05)
- There was a positive correlation between motivation and engagement
- ($r$=.81, $p$=<.01)
- Test for homogeneity of regression slopes showed the assumption was not violated ($p$>.05)

These show that assumptions for ANCOVA were met for engagement. Therefore, ANCOVA was used for engagement as in satisfaction

Then, the main effect of video type (linear or interactive) based on estimated marginal means was compared in terms of engagement. Table 25 presents the comparison results for engagement when motivation was controlled:

**Table 25.** Comparison of Mean Scores for Engagement (when motivation was controlled)

| Video Type | *n* (Observation) | Mean | *SD* | Mean (adjusted) | *SE* | *F* | eta$^2$ |
|---|---|---|---|---|---|---|---|
| **Linear** | 44 | 3.41 | 1.24 | 3.19 | 0.11 | 0.076* | 0.001 |
| **Interactive** | 53 | 3.00 | 1.19 | 3.15 | 0.10 | | |

*$p$ = .784 > .05

The results show that no significant difference was found between the linear ($M$ = 3.19, $SD$ = 0.11) and interactive videos ($M$ = 3.15, $SD$ = 0.10) in terms of students' engagement scores when motivation was controlled, $p$ > .05.

### 4.1.2.3 Learning Performance

*Hypothesis 6:* There is a significant difference between linear and interactive videos in terms of students' learning performance when motivation is controlled.

*Null hypothesis:* There is no significant difference between linear and interactive videos in terms of students' learning performance when motivation is controlled.

The assumptions of ANCOVA for learning performance were firstly controlled. Here are the results for the assumptions in terms of learning performance:

- Kolmogorov-Smirnov test showed normality of residuals was met ($p$>.05).
- Levene's test indicated variances are equal ($p$<.05)
- There was a positive correlation between motivation and learning performance
- ($r$=.25, $p$=<.01)
- Test for homogeneity of regression slopes showed the assumption was not violated ($p$>.05)

Homogeneity of variance was violated as can be seen from the list above. Therefore, as a nonparametic test, Quade's ANCOVA was used for learning performance. Rank of the learning performance scores were compared in Quade's ANCOVA. Table 26 shows the comparison results for learning performance scores when motivation was controlled:

**Table 26.** Comparison of Mean Scores for Learning performance (when motivation was controlled)

| Video Type | *n* (observation) | Mean | *SD* | Mean (adjusted rank) | *SE* | *F* | eta$^2$ |
|---|---|---|---|---|---|---|---|
| **Linear** | 44 | 85.40 | 18.66 | -2.20 | 2.31 | 1.304* | 0.021 |
| **Interactive** | 53 | 84.62 | 14.35 | 1.31 | 2.02 | | |

*$p$ = .258 > .05

The results show that no significant difference was found between the linear ($M_{rank}$ = -2.20, $SD_{rank}$ = 2.31) and interactive videos ($M_{rank}$ = 1.31, $SD_{rank}$ = 2.02) in terms of students' learning performance when motivation was controlled, $p$ > .05.

Table 27 presents a summary of the findings regarding all hypotheses for both the first and second research questions.

**Table 27.** Summary Findings for RQ1 and RQ2

| *No.* | *Hypothesis* | **Students/ Observations* | *Statistical technique* | *Mean (Lin.)* | *Mean (Int.)* | *p* | *Accepted* |
|---|---|---|---|---|---|---|---|
| 1 | There is a significant difference between linear and interactive videos in terms of students' satisfaction. | 18 students | *t*-test | 2.54 | 2.50 | .782 | No |
| 2 | There is a significant difference between linear and interactive videos in terms of students' engagement. | 18 students | *t*-test | 3.40 | 3.00 | .070 | No |
| 3 | There is a significant difference between linear and interactive videos in terms of students' learning performance. | 18 students | *t*-test | 84.56 | 84.38 | .948 | No |
| 4 | There is a significant difference between linear and interactive videos in terms of students' satisfaction score when motivation is controlled. | 97 obs. | ANCOVA | 2.43 | 2.59 | .039 | Yes |
| 5 | There is a significant difference between linear and interactive videos in terms of students' engagement score when motivation is controlled. | 97 obs. | ANCOVA | 3.19 | 3.15 | .784 | No |
| 6 | There is a significant difference between linear and interactive videos in terms of students' learning performance when motivation is controlled. | 97 obs. | Quade's ANCOVA | -2.2** | 1.31** | .258 | No |

* 97 observations of 18 students (44 from linear videos, 53 from interactive videos)
** Rank was used for Quade's ANCOVA

### 4.1.3 Differences Based on Click Reasons

*RQ3:* Are there any significant differences between linear and interactive videos in terms of students' click reasons?

The third research question includes a total of seven hypotheses (H7, H8, H9, H10, H11, H12, H13) since seven different click reasons were found to require intervention (see Figure 5). This research question aims to address changes in click reason frequency according to the framework.

*Hypothesis 7:* There is a significant difference in the frequency of students' notetaking between linear and interactive videos.

*Null hypothesis:* There is no significant difference in the frequency of students' notetaking between linear and interactive videos.

The Shapiro-Wilk test results for the frequency of students' notetaking showed that the measured differences were normally distributed ($p > .05$). The frequency of students' notetaking between the linear and interactive videos was compared with a paired samples *t*-test. Table 28 illustrates the comparison results for notetaking:

**Table 28.** Comparison of Notetaking between Linear and Interactive Videos

| Video Type | *n* | $\bar{X}$ | *SD* | *df* | *t* | *p* |
|---|---|---|---|---|---|---|
| **Linear** | 17 | 2.76 | 1.42 | 16 | 0.389 | .702 |
| **Interactive** | 17 | 2.71 | 1.45 | | | |

The results show that no significant difference was found between the linear ($M = 2.76$, $S = 1.42$) and interactive videos ($M = 2.71$, $SD = 1.45$) in terms of the students' frequency of notetaking, $t(16) = 0.389$, $p > .05$.

*Hypothesis 8:* There is a significant difference in the frequency of students' losing concentration between linear and interactive videos.

*Null hypothesis:* There is no significant difference in the frequency of students' losing concentration between linear and interactive videos.

The Shapiro-Wilk test results for the frequency of students' losing concentration showed that the measured differences were normally distributed ($p > .05$). The frequency of students' losing concentration between the linear and interactive videos was compared with a paired samples *t*-test. Table 29 presents the comparison results for losing concentration:

**Table 29.** Comparison of Losing Concentration between Linear and Interactive Videos

| Video Type | *n* | $\bar{X}$ | *SD* | *df* | *t* | *p* |
|---|---|---|---|---|---|---|
| **Linear** | 18 | 2.71 | 0.95 | 17 | 0.553 | .588 |
| **Interactive** | 18 | 2.61 | 0.85 | | | |

The results show that no significant difference was found between the linear ($M$ = 2.71, $S$ = 0.95) and interactive videos ($M$ = 2.61, $SD$ = 0.85) in terms of the students' frequency of losing concentration, $t(17) = 0.55$, $p > .05$.

*Hypothesis 9:* There is a significant difference in the frequency of students' not understanding the content between linear and interactive videos.

*Null hypothesis:* There is no significant difference in the frequency of students' not understanding the content between linear and interactive videos.

The Shapiro-Wilk test results for the frequency of students' not understanding the content showed that the measured differences were normally distributed ($p > .05$). The frequency of students' not understanding the content between linear and interactive videos was compared with a paired samples *t*-test. Table 30 shows the comparison results for not understanding the content:

**Table 30.** Comparison of Not Understanding Content between Linear and Interactive Videos

| Video Type | *n* | $\bar{X}$ | *SD* | *df* | *t* | *p* |
|---|---|---|---|---|---|---|
| **Linear** | 18 | 1.83 | 0.83 | 17 | -0.408 | .689 |
| **Interactive** | 18 | 1.90 | 0.79 | | | |

The results show that no significant difference was found between the linear ($M$ = 1.83, $S$ = 0.83) and interactive videos ($M$ = 1.89, $SD$ = 0.79) in terms of the students' frequency of not understanding the content, $t(17) = -0.41$, $p > .05$.

*Hypothesis 10:* There is a significant difference in the frequency of students' going back to review important points between linear and interactive videos.

*Null hypothesis:* There is no significant difference in the frequency of students' going back to review important points between linear and interactive videos.

The Shapiro-Wilk test results for the frequency of students' going back to review important points showed that the measured differences were normally distributed ($p > .05$). The frequency of students' going back to review important points between the linear

and interactive videos was compared with a paired samples *t*-test. Table 31 illustrates the comparison results for going back to review important points:

**Table 31.** Comparison of Going Back to Review Important Points between Linear and Interactive Videos

| Video Type | *n* | $\bar{X}$ | *SD* | *df* | *t* | *p* |
|---|---|---|---|---|---|---|
| **Linear** | 18 | 2.97 | 1.13 | 17 | 2.742 | .014 |
| **Interactive** | 18 | 2.57 | 1.08 | | | |

The results show that a significant difference was found between the linear ($M$ = 2.97, $S$ = 1.13) and interactive videos ($M$ = 2.57, $SD$ = 1.08) in terms of the students' frequency of going back to review important points, $t(17) = 2.74$, $p < .05$.

*Hypothesis 11:* There is a significant difference in the frequency of students' skipping unimportant points between linear and interactive videos.

*Null hypothesis:* There is no significant difference in the frequency of students' skipping unimportant points between linear and interactive videos.

The Shapiro-Wilk test results for the frequency of students' skipping unimportant points showed that the measured differences were normally distributed ($p > .05$). The frequency of students' skipping unimportant points between the linear and interactive videos was compared with a paired samples *t*-test. Table 32 presents the comparison results for skipping unimportant points:

**Table 32.** Comparison of Skipping Unimportant Points between Linear and Interactive Videos

| Video Type | *n* | $\bar{X}$ | *SD* | *df* | *t* | *p* |
|---|---|---|---|---|---|---|
| **Linear** | 18 | 1.54 | 0.69 | 17 | -3.292 | .004 |
| **Interactive** | 18 | 2.10 | 0.82 | | | |

The results show that a significant difference was found between the linear ($M$ = 1.54, $S$ = 0.69) and interactive videos ($M$ = 2.10, $SD$ = 0.82) in terms of the students' frequency of skipping unimportant points, $t(17) = -3.29$, $p < .05$.

*Hypothesis 12:* There is a significant difference in the frequency of students' wanting to finish quickly between linear and interactive videos.

*Null hypothesis:* There is no significant difference in the frequency of students' wanting to finish quickly between linear and interactive videos.

The Shapiro-Wilk test results for the frequency of students' wanting to finish quickly showed that the measured differences were normally distributed ($p > .05$). The frequency of students' wanting to finish quickly between the linear and interactive videos was compared with a paired samples *t*-test. Table 34 presents the comparison results for students' wanting to finish quickly:

**Table 33.** Comparison of Wanting to Finish Quickly between Linear and Interactive Videos

| Video Type | *n* | $\bar{X}$ | *SD* | *df* | *t* | *p* |
|---|---|---|---|---|---|---|
| **Linear** | 18 | 3.10 | 1.27 | 17 | -0.367 | .718 |
| **Interactive** | 18 | 3.17 | 1.12 | | | |

The results show that no significant difference was found between the linear ($M = 3.10$, $S = 1.27$) and interactive videos ($M = 3.17$, $SD = 1.12$) in terms of the students' frequency of wanting to finish quickly, $t(17) = -0.367$, $p > .05$.

*Hypothesis 13:* There is a significant difference in the frequency of students' facing language issues between linear and interactive videos.

*Null hypothesis:* There is no significant difference in the frequency of students' facing language issues between linear and interactive videos.

The Shapiro-Wilk test results for the frequency of students being faced with language issues showed that the measured differences were not normally distributed ($p < .05$). The frequency of students' facing language issues between linear and interactive videos was compared with the Wilcoxon signed-rank test since it did not meet the assumptions for a *t*-test. Table 31 presents the comparison results for students' facing language issues:

**Table 34.** Comparison of Facing Language Issues between Linear and Interactive Videos

| Video Type | *n* | $\bar{X}$ | *SD* | *Z* | *p* |
|---|---|---|---|---|---|
| **Linear** | 16 | 1.93 | 0.92 | 1.615 | .106 |
| **Interactive** | 16 | 1.78 | 0.88 | | |

The results show that no significant difference was found between the linear ($M = 1.93$, $S = 0.92$) and interactive videos ($M = 1.78$, $SD = 0.88$) in terms of the frequency of students being faced with language issues, $Z = 1.615$, $p > .05$.

Since all of the reasons were investigated within the same section in the questionnaire and there were multiple comparisons performed, *p*-values were also adjusted according to the Benjamini-Hochberg procedure (Benjamini & Hochberg, 1995). Table 35 presents a summary, together with an adjusted *p*-value.

**Table 35.** Comparison of Reasons between Linear and Interactive Videos

| *Hypothesis* | *Reason* | *Video Type* | *n* | *M* | *SD* | *df* | *t* | *p (adjusted)* | *Accepted* |
|---|---|---|---|---|---|---|---|---|---|
| **7** | Notetaking | Linear | 17 | 2.76 | 1.42 | 16 | 0.389 | .718 | No |
| | | Interactive | 17 | 2.71 | 1.45 | | | | |
| **8** | Losing Concentration | Linear | 18 | 2.71 | 0.95 | 17 | 0.553 | .718 | No |
| | | Interactive | 18 | 2.61 | 0.85 | | | | |
| **9** | Lack of Understanding Content | Linear | 18 | 1.83 | 0.83 | 17 | -0.408 | .718 | No |
| | | Interactive | 18 | 1.90 | 0.79 | | | | |
| **10** | Going Back to Review Important Points | Linear | 18 | 2.97 | 1.13 | 17 | 2.742 | .049 | Yes |
| | | Interactive | 18 | 2.57 | 1.08 | | | | |
| **11** | Skipping Unimportant Points | Linear | 18 | 1.54 | 0.69 | 17 | -3.292 | .028 | Yes |
| | | Interactive | 18 | 2.10 | 0.82 | | | | |
| **12** | Wanting to Finish Quickly | Linear | 18 | 3.10 | 1.27 | 17 | -0.367 | .718 | No |
| | | Interactive | 18 | 3.17 | 1.12 | | | | |
| ***Hypothesis*** | ***Reason*** | ***Video Type*** | ***n*** | ***M*** | ***SD*** | | ***Z** | ***p (adjusted)*** | ***Accepted*** |
| **13** | Facing Language Issues | Linear | 16 | 1.93 | 0.92 | | -1.615 | .247 | No |
| | | Interactive | 16 | 1.78 | 0.88 | | | | |

*Wilcoxon signed-rank test was used for H13 since it did not meet assumptions for *t*-test.

The results presented in Table 35 show that significant differences were found to exist for two out of the seven click reasons ("Going Back to Important Points" and "Skipping Unimportant Points"). First, a significant difference was found to exist between the linear ($M = 2.97$, $SD = 1.13$) and interactive videos ($M = 2.57$, $SD = 1.08$) in terms of the students' frequency of going back to view important points ($t(17) = 2.74$, $p < .05$). Second, a significant difference was found to exist between the linear ($M = 1.54$, $SD = 0.69$) and interactive videos ($M = 2.10$, $SD = 0.82$) in terms of the students' frequency of skipping unimportant points ($t(17) = -3.29$, $p < .05$).

In addition, the students were shown to have taken notes more ($M_{linear} = 2.76 > M_{interactive} = 2.71$), lost concentration more ($M_{linear} = 2.71 > M_{interactive} = 2.61$), and faced language issues more ($M_{linear} = 1.93 > M_{interactive} = 1.78$) in the linear videos than in the interactive videos, even where the results were found to be non-significant.

### 4.2 Results for Comprehensive Learning Analytics

Correlations, clusters, and predictors based on the participant students' video clickstream behaviors were investigated with comprehensive learning analytics in order to address the last three research questions.

The fourth research question focuses on correlations between the students' video clickstream behaviors and their learning performance, whilst the fifth research question aims to find clusters based on the students' video clickstream behaviors. The sixth research question then aims to reveal the predictors of engagement, satisfaction, and learning performance and to examine how successfully these predictors can classify engagement, satisfaction, and learning performance.

Since the video clickstream behaviors were collected during the viewing of the first three videos, the vieving behavior results relate only to the linear videos. Analysis of the video clickstream behaviors was based on repeated measures instead of single aggregated measure in this section because students may well perform different behaviors when viewing different videos. Therefore, data collected representing 16 of the participant students who watched more than one video lecture were used to analyze the video clickstream behaviors. That is, data for the two students who only viewed one of the video lectures were eliminated from the watching behaviors analysis (which differs to the analysis of the framework). While those students' data included valuable information for the evaluation of the framework as they had watched all three of the interactive videos, their data was insufficient for the purposes of investigating video clickstream behaviors.

In total, 42 observations were obtained from 16 participant students in the experiment. Table 36 presents the numbers of interactions for each participant based on each video.

**Table 36.** Number of Interactions for Participants per Video Lecture

| | *Video Lecture 1* | | | | | *Video Lecture 2* | | | | | *Video Lecture 3* | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | ***NPl*** | ***NP*** | ***NF*** | ***NB*** | ***T*** | ***NPl*** | ***NP*** | ***NF*** | ***NB*** | ***T*** | ***NPl*** | ***NP*** | ***NF*** | ***NB*** | ***T*** | ***Total*** |
| P01 | 35 | 34 | 1 | 12 | 89 | 9 | 7 | 0 | 5 | 29 | 53 | 52 | 0 | 6 | 118 | 236 |
| P02 | | | - | | | 3 | 2 | 0 | 1 | 13 | 2 | 1 | 0 | 0 | 10 | 23 |
| P03 | 2 | 1 | 0 | 0 | 5 | | | - | | | 2 | 1 | 5 | 0 | 15 | 20 |
| P04 | 4 | 3 | 0 | 1 | 11 | 6 | 6 | 3 | 4 | 31 | 24 | 23 | 0 | 8 | 65 | 107 |
| P05 | | | - | | | 11 | 11 | 5 | 15 | 50 | 16 | 15 | 15 | 17 | 68 | 118 |
| P06 | 22 | 17 | 32 | 30 | 127 | 20 | 18 | 17 | 41 | 104 | 24 | 23 | 6 | 26 | 86 | 317 |
| P07 | 3 | 0 | 0 | 0 | 10 | 6 | 5 | 10 | 0 | 30 | 1 | 0 | 0 | 0 | 2 | 42 |
| P08 | 3 | 0 | 0 | 0 | 11 | 4 | 0 | 0 | 0 | 11 | 33 | 31 | 3 | 22 | 100 | 122 |
| P09 | 16 | 16 | 2 | 11 | 48 | 8 | 7 | 1 | 3 | 25 | | | - | | | 73 |
| P10 | | | - | | | 3 | 2 | 7 | 0 | 16 | 2 | 2 | 8 | 1 | 15 | 31 |
| P11 | 67 | 67 | 6 | 13 | 160 | 64 | 63 | 17 | 46 | 198 | 3 | 2 | 0 | 0 | 8 | 366 |
| P12 | 49 | 48 | 1 | 9 | 117 | | | - | | | 7 | 6 | 15 | 1 | 32 | 149 |
| P13 | 9 | 6 | 3 | 20 | 46 | 3 | 2 | 0 | 2 | 9 | 6 | 4 | 0 | 3 | 21 | 76 |
| P14 | 9 | 8 | 7 | 9 | 38 | 9 | 8 | 0 | 3 | 23 | 14 | 13 | 17 | 11 | 58 | 119 |
| P15 | 1 | 1 | 0 | 1 | 5 | 18 | 18 | 7 | 12 | 58 | 11 | 11 | 3 | 7 | 40 | 103 |
| P16 | 4 | 3 | 0 | 0 | 11 | 4 | 4 | 0 | 1 | 11 | 8 | 7 | 1 | 0 | 28 | 50 |
| ***Total*** | 228 | 206 | 52 | 106 | 692 | 151 | 138 | 55 | 117 | 529 | 186 | 173 | 50 | 84 | 573 | 1,794 |

*NPl = # of Plays, NP = # of Pauses, NB = # of Backwards, NF = # of Forwards, T = Total Interactions (incl. other types). Rows show each participant's interaction statistics for each video lecture.*

Since *number of pauses* and *number of plays* are highly correlated, only *number of pauses* was used in the analyses to answer all hypotheses except hypothesis 14 (H14 investigates *total number of interactions*).

### 4.2.1 Correlations Between Video Clickstream Behaviors and Learning Performance

*RQ4:* Are there any correlations between video clickstream behaviors and learning performance?

The fourth research question includes six hypotheses (H14, H15, H16, H17, H18, H19) since it aims to investigate the effect of the *total number of clicks*, *number of pauses*, *number of backwards*, *number of forwards*, *forward speed*, and *backward speed* separately on the students' learning performance.

The "rmcorr: Repeated Measures Correlation" package was used in R to calculate the correlations between the students' video clickstream behaviors and their learning performance. The results and hypotheses for the fourth research question are detailed in the following paragraphs. In addition, Table 37 presents a summary of the correlation findings at the end of this section.

*Hypothesis 14:* There is a significant correlation between *total number of clicks* and learning performance.

*Null hypothesis:* There is no significant correlation between *total number of clicks* and learning performance.

The results show that a significant positive correlation was found between *total number of interactions* and learning performance ($n = 42$, $r = .66$, $p < .01$).

Since *number of pauses* and *number of plays* are highly correlated as in Figure 3, using both to calculate the *total number of interactions* can mislead. However, all data are almost affected by this, which does not have a significant effect on this hypothesis. To show this, the *total number of interactions* was also calculated by subtracting *number of plays* from the total and repeated measure analysis gave almost same result with previous one ($n = 42$, $r = .65$, $p < .01$).

*Hypothesis 15:* There is a significant correlation between *number of pauses* and learning performance.

*Null hypothesis:* There is no significant correlation between *number of pauses* and learning performance.

The results show that a significant positive correlation was found between *number of pauses* and learning performance ($n = 42$, $r = .62$, $p < .01$).

*Hypothesis 16:* There is a significant correlation between *number of backwards* and learning performance.

*Null hypothesis:* There is no significant correlation between *number of backwards* and learning performance.

The results show that a significant positive correlation was found between *number of backwards* and learning performance ($n = 42$, $r = .48$, $p = .01$).

*Hypothesis 17:* There is a significant correlation between *number of forwards* and learning performance.

*Null hypothesis:* There is no significant correlation between *number of forwards* and learning performance.

The results show that a significant positive correlation was found between the *number of forwards* and learning performance ($n = 42$, $r = .32$, $p = .10$).

The *backward speed* and *forward speed* for each interaction was also calculated by dividing skipping time by the video total time, as follows:

$$BS_{i,j} = BT_{i,j} / VT_{i,j}$$

$$FS_{i,j} = FT_{i,j} / VT_{i,j}$$ where $i$ = related user *and* $j$ = related video lecture

*BS = Backward Speed, BT = Skipping Time for Backwards (Mean), FT = Skipping Time for Forwards (Mean), VT = Video Time (Total)*

*Hypothesis 18:* There is a significant correlation between the *mean of backward speed* and learning performance.

*Null hypothesis:* There is no significant correlation between the *mean of backward speed* and learning performance.

The results show that a significant negative correlation was found between the *mean of backward speed* and learning performance ($r = -.65$, $p < .01$).

*Hypothesis 19:* There is a significant correlation between the *mean of forward speed* and learning performance.

*Null hypothesis:* There is no significant correlation between the *mean of forward* speed and learning performance.

The results show that no significant correlation was found between the *mean of forward speed* and learning performance ($r = -.28$, $p > .05$).

**Table 37.** Summary of Correlation Findings

| *No* | *Hypothesis* | *r* | *p* | *Accepted* |
|---|---|---|---|---|
| 14 | There is a significant correlation between total number of clicks and learning performance. | .66 | < .01 | Yes |
| 15 | There is a significant correlation between number of pauses and learning performance. | .62 | < .01 | Yes |
| 16 | There is a significant correlation between number of backwards and learning performance. | .48 | = .01 | Yes |
| 17 | There is a significant correlation between number of forwards and learning performance. | .32 | > .05 | No |
| 18 | There is a significant correlation between the mean of backward speed and learning performance. | .65 | < .01 | Yes |
| 19 | There is a significant correlation between the mean of forward speed and learning performance. | .28 | > .05 | No |

### 4.2.2 Clusters Based on Video Clickstream Behaviors

*RQ5:* Are there any significant differences between clusters within cluster sets based on video clickstream behaviors in terms of learning performance?

The fifth research question also includes three hypotheses (H20, H21, H22), since three clustering types were determined in order to answer the research question. Details for these clusters are explained in detail in the following paragraph. At the end of this section, Table 45 presents a summary of all the clustering findings.

The fifth research question examined the differences between behavioral groups obtained based on the students' video clickstream interactions. First, the groups were identified using cluster analysis and their differences were then investigated using statistical techniques. The students' clickstream behaviors were clustered using three different feature sets according to the *two-step cluster analysis* approach, which is considered preferable for this type of analysis since it merges the principles of traditional clustering procedures and presents an overall goodness of fit measure (Polier, 2019). Cluster graphs were also used in order to verify the quality of the clusters. The following lists the selections for the two-step cluster analysis:

1. Log-likelihood selected as Distance Measure
2. Schwarz's Bayesian Criterion selected as Clustering Criterion

The feature sets used for the analysis are as follows:

*Feature Set 1: Numbers of pauses, numbers of forwards,* and *numbers of backwards* for each user for each video

*Feature Set 2: Percentage of pauses, percentage of forwards,* and *percentage of backwards* for each user for each video

*Feature Set 3: Forward speed* and *backward speed* for each user for each video

As a result, the students' video clickstream behaviors were clustered in three different ways:

1. Clustering based on Numbers of Pauses, Forwards, and Backwards
2. Clustering based on Percentage of Pauses, Forwards, and Backwards
3. Clustering based on Forward Speed and Backward Speed

Therefore, three clustering sets were created according to the three aforementioned feature sets used for the analysis.

#### 4.2.2.1 Clustering Based on Numbers of Pauses, Forwards, Backwards

Figure 6 shows the number of clusters and Schwarz's Bayesian Criterion (BIC). Two was selected as the number of clusters since it shows the smallest BIC value.

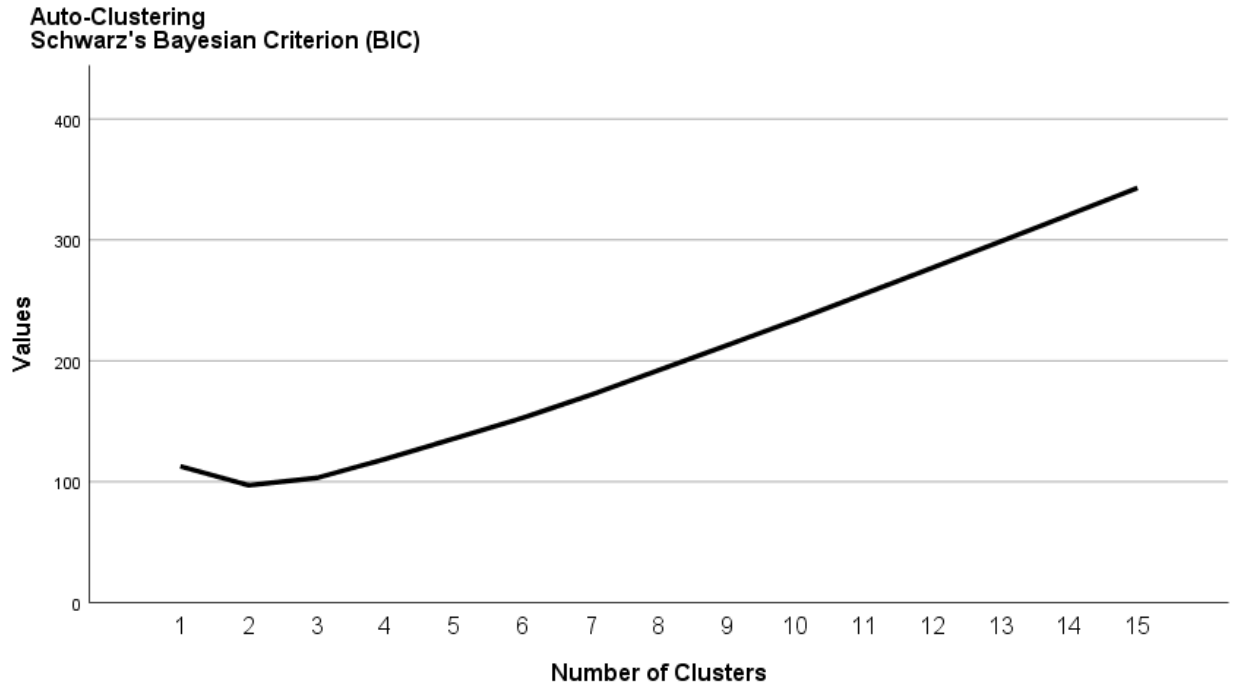



**Figure 6.** BIC Values for Number of Clusters Based on Number of Clicks

Table 38 shows the number of *pauses*, *backwards* and *forwards* interactions for each cluster. Clusters were named as (1) "High Clicking" and (2) "Low Clicking" based on the mean scores of the click numbers.

**Table 38.** Clusters Based on Numbers of Pauses, Forwards and Backwards

| | | | *Pause* | | *Backward* | | *Forward* | |
|---|---|---|---|---|---|---|---|---|
| | | ***N*** | ***M*** | ***SD*** | ***M*** | ***SD*** | ***M*** | ***SD*** |
| **Cluster** | **High Clicking** | 14 | 29.71 | 20.07 | 20.00 | 11.98 | 9.29 | 9.09 |
| | **Low Clicking** | 28 | 4.71 | 5.20 | 2.18 | 3.10 | 2.21 | 3.84 |
| | **Combined** | | 13.05 | 16.97 | 8.12 | 11.14 | 4.57 | 6.88 |

Figure 7 shows the average Silhouette value for the clustering. As can be seen, the clustering achieved a value of 0.6, which is considered as "good."




**Figure 7.** Cluster Quality Based on Numbers of Clicks

Figure 8 shows the distribution of the clusters based on *numbers of pauses, forwards,* and *backwards.*

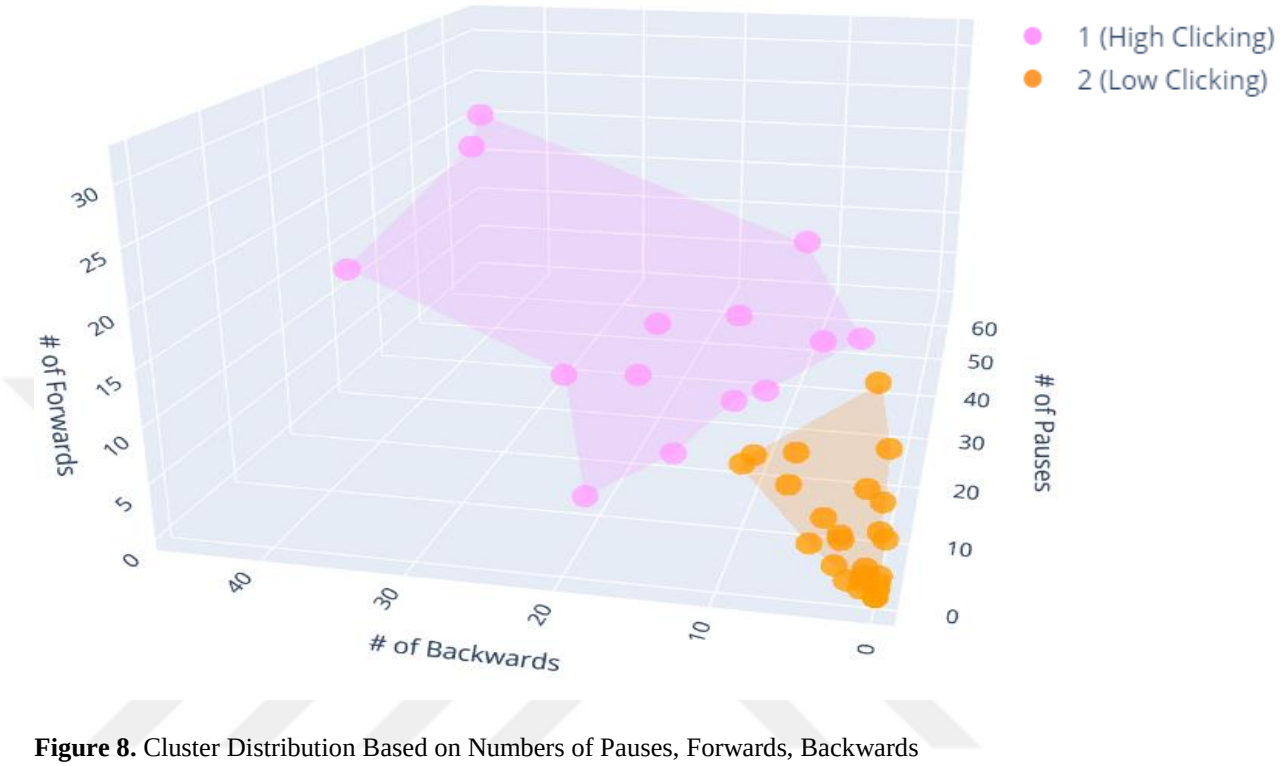



**Figure 8.** Cluster Distribution Based on Numbers of Pauses, Forwards, Backwards

*Hypothesis 20:* There is a significant difference in learning performance between the groups based on the number of clicks.

*Null hypothesis:* There is no significant difference in learning performance between the groups based on the number of clicks.

Table 39 presents the mean learning performance scores for each group.

**Table 39.** Robust Linear Mixed-Model Results to Predict Learning Performance for Clusters Based on Numbers

| ***Response*** | ***Predictor*** | ***Estimate*** | ***SE*** | ***df*** | ***t*** | ***p*** |
|---|---|---|---|---|---|---|
| **Learning performance** | (Intercept) | 81.06 | 3.04 | 40 | 26.70 | .000*** |
| | Cluster (Ref = Low Clicking) High Clicking | 18.23 | 5.26 | 40 | 3.47 | .001*** |

The learning performance was predicted using the two clustering groups in order to investigate Hypothesis 20 using the Robust Linear Mixed-Model Analysis technique (Koller, 2016). As an individual student could belong to either the low-clicking or high-clicking clusters for a certain video, the points in each group might not be treatable as independent and residuals may be non-normal based on categorical data. Hence, the

Robust Linear Mixed-Effect Model (Robust LMM) was employed, which is a linear additive model of multi-levels by groups, where the random effect within group is additive to the fixed effect between group:

$$Y_i = X_i\beta + Z_i b_i + \varepsilon_i$$

where $Y$ is the vector of the response variable (quiz scores), $X$ is the fixed effect (video interaction features), $\beta$ is the vector of the fixed effects coefficients, $Z$ is the random effects (each user), $b$ is the vector of random effects coefficients, $i$ is the related group, and $\varepsilon$ is the error for random effects (Fox, 2002).

Robust LMM does not make any assumption, with one exception being that the model parameters can be estimated as mentioned in section 3.5.

**Table 40.** Robust Linear Mixed Model Results to Predict Learning Performance for Clusters Based on Numbers

| ***Response*** | ***Predictor*** | ***Estimate*** | ***SE*** | ***df*** | ***t*** | ***p*** |
|---|---|---|---|---|---|---|
| **Learning performance** | (Intercept) | 81.06 | 3.04 | 40 | 26.70 | .000*** |
| | Cluster (Ref = Low Clicking) High Clicking | 18.23 | 5.26 | 40 | 3.47 | .001*** |

A statistically significant group effect ($p$ = .001) was found, which means that high-clicking behaviors resulted in higher learning performance when compared to low-clicking behaviors (Estimate = +18.23).

#### 4.2.2.2 Clustering Based on Percentage of Pauses, Forwards, Backwards

For each participant student per video, the *Percentages of Pauses*, *Forwards*, and *Backwards* were calculated by dividing the number of interactions for each related click type by the total number of interactions as follows:

$$PP_{i,j} = NP_{i,j} / (NP_{i,j}+NB_{i,j}+NF_{i,j})$$

$$PB_{i,j} = NB_{i,j} / (NP_{i,j}+NB_{i,j}+NF_{i,j})$$

$$PP_{i,j} = NF_{i,j} / (NP_{i,j}+NB_{i,j}+NF_{i,j})$$

*PP = % of Pauses, PB = % of Backwards, PF = % of Forwards, NP = # of Pauses, NB = # of Backwards, NF = # of Forwards*

The two-step cluster produced three clusters, one of which is a combination of the forwards with those not having all three basic interaction types (*pause*, *forward*, *backward*). Therefore, in order to obtain clusters based on *percentage of pauses, forwards,*

and *backwards*, hierarchical clustering based on average linkage between groups was applied. Four was selected as the number of clusters since it was the most suited to the group based on percentage when the data and dendrogram were analyzed (see Appendix H).

Table 41 shows the *percentage of pause*, *backward* and *forward* for each cluster. The mean scores show that first cluster had the highest *percentage of pauses*, the second cluster had the highest *percentage of backwards*, whilst the third cluster had the highest *percentage of forwards*. Therefore, the clusters can be named as (1) *Pauser*, (2) *Backwarder*, (3) *Forwarder*, and (4) *Player* from looking at the mean click percentage scores.

**Table 41.** Clusters Based on Percentage of Pauses, Forwards and Backwards

| | | ***N*** | ***Pause %*** | | ***Forward %*** | | ***Backward %*** | |
|---|---|---|---|---|---|---|---|---|
| | | | ***M*** | ***SD*** | ***M*** | ***SD*** | ***M*** | ***SD*** |
| **Cluster** | **Pauser** | 15 | .83 | .13 | .02 | .04 | .15 | .12 |
| | **Backwarder** | 18 | .42 | .13 | .15 | .13 | .42 | .10 |
| | **Forwarder** | 5 | .24 | .07 | .74 | .07 | .03 | .04 |
| | **Player** | 4 | .00 | .00 | .00 | .00 | .00 | .00 |
| | **Combined** | | .51 | .30 | .16 | .24 | .24 | .19 |

Figure 9 shows the cluster distribution based on *Percentage of Pauses*, *Forwards,* and *Backwards*.

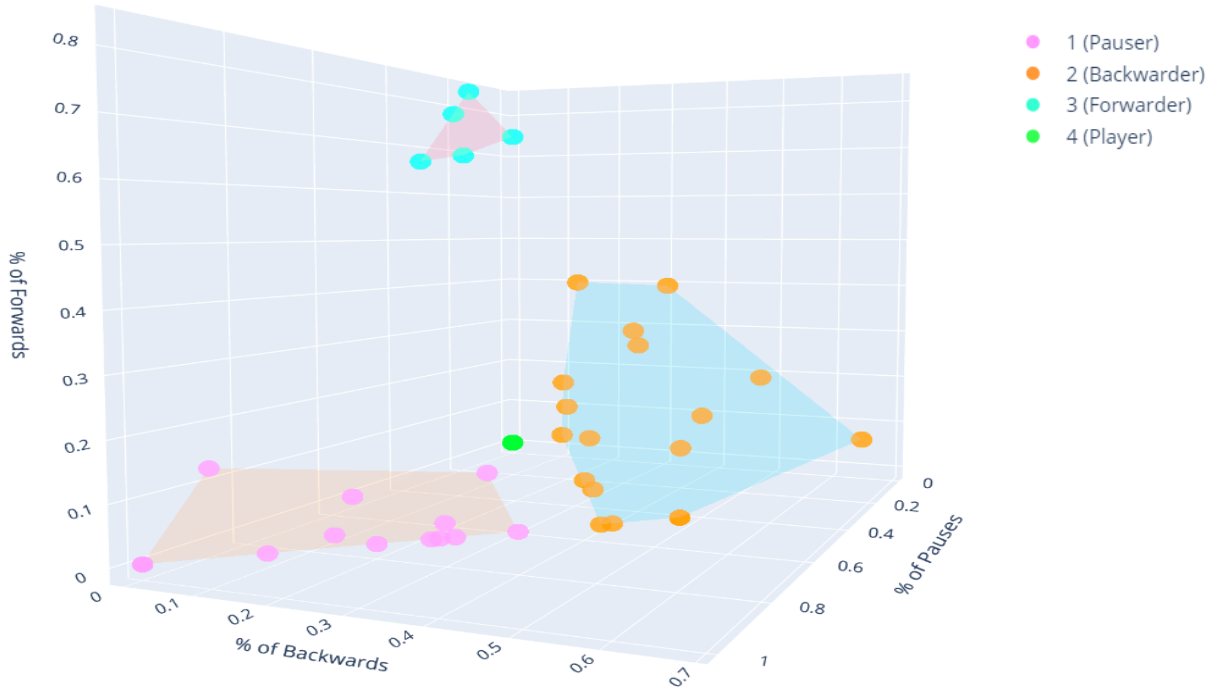



**Figure 9.** Cluster Distribution Based on Percentage of Pauses, Forwards, Backwards

*Hypothesis 21:* There is a significant difference in learning performance between the groups based on their click preferences.

*Null hypothesis:* There is no significant difference in learning performance between the groups based on their click preferences.

Next, the learning performance differences between the groups based on their click preferences were compared with those of the statistical analyses. For this, the Robust Linear Mixed-Model approach was employed since it considers the randomized effects of repeated measures differently from both ANOVA and Kruskal Wallis.

**Table 42.** Robust Linear Mixed Model Results to Predict Learning Performance for Clusters Based on Percentage

| ***Response*** | ***Predictor*** | ***Estimate*** | ***SE*** | ***df*** | ***t*** | ***p*** |
|---|---|---|---|---|---|---|
| **Learning performance** | (Intercept) | 71.45 | 8.85 | 38 | 8.07 | .000*** |
| | Cluster (Ref = Player) Pauser | 14.01 | 9.96 | 38 | 1.41 | .168 |
| | Cluster (Ref = Player) Backwarder | 22.03 | 9.79 | 38 | 2.25 | .030* |
| | Cluster (Ref = Player) Forwarder | 11.14 | 11.88 | 38 | 0.94 | .354 |

The results show that the clusters based on *Percentage of Pauses*, *Forwards*, and *Backwards* significantly predict learning performance ($p$ = .030). The *Backwarder* results produced a higher learning performance when compared to *Player* (Estimate = +22.03).

#### 4.2.2.3 Clustering Based on Forward Speed and Backward Speed

Figure 10 shows the number of clusters and Schwarz's Bayesian Criterion (BIC). Two was selected as the number of clusters since it shows the smallest BIC value.

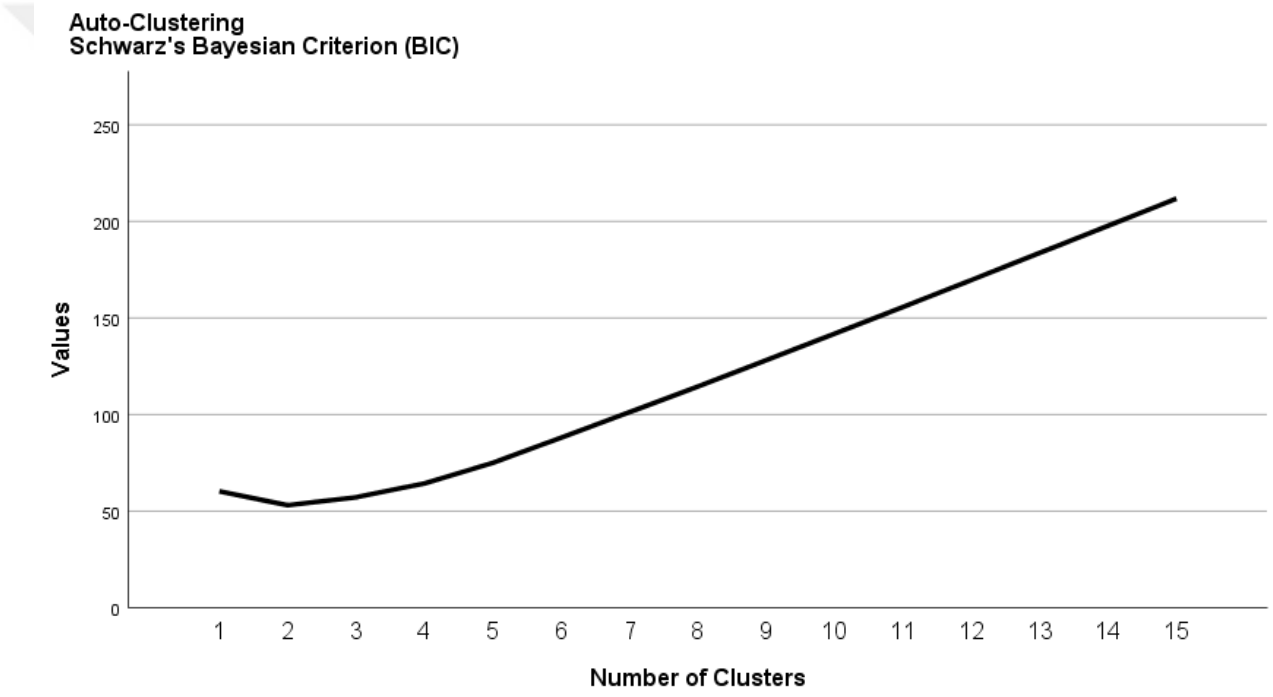



**Figure 10.** BIC Values for Number of Clusters Based on Click Speed

Table 43 shows the means of *backward speed* and *forward speed* for each cluster. The mean scores show that first cluster has high scores whilst the second has low scores for both *forward speed* and *backward speed*. Therefore, these clusters are named as (1) "Fast Watching" and (2) "Slow Watching" in accordance with their respective mean *forward speed* and *backward speed* scores.

**Table 43.** Clusters Based on Forward Speed and Backward Speed

| | | *N* | *Forward Speed* | | *Backward Speed* | |
|---|---|---|---|---|---|---|
| | | | ***M*** | ***SD*** | ***M*** | ***SD*** |
| **Cluster** | **Fast Watching** | 8 | .083 | .066 | .071 | .084 |
| | **Slow Watching** | 25 | .008 | .008 | .016 | .016 |
| | **Combined** | | .027 | .046 | .030 | .048 |

Figure 11 shows the average Silhouette value for clustering. As can be seen, the clustering has a value of 0.7, which is considered to be "good."

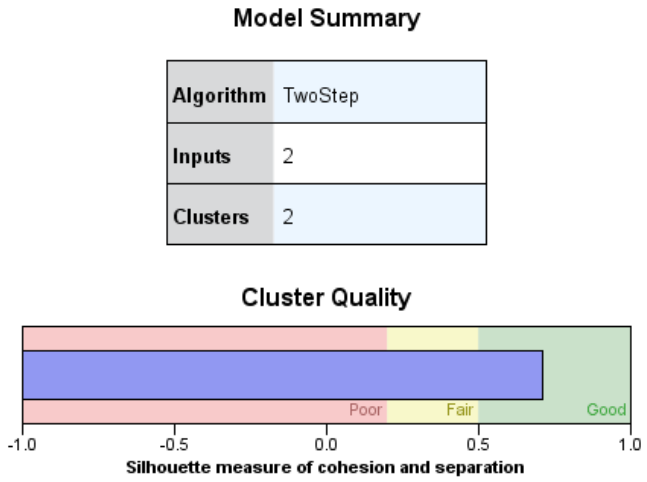



**Figure 11.** Cluster Quality (Based on Click Speed)

Figure 12 shows the distribution of clusters based on *forward speed* and *backward speed*.

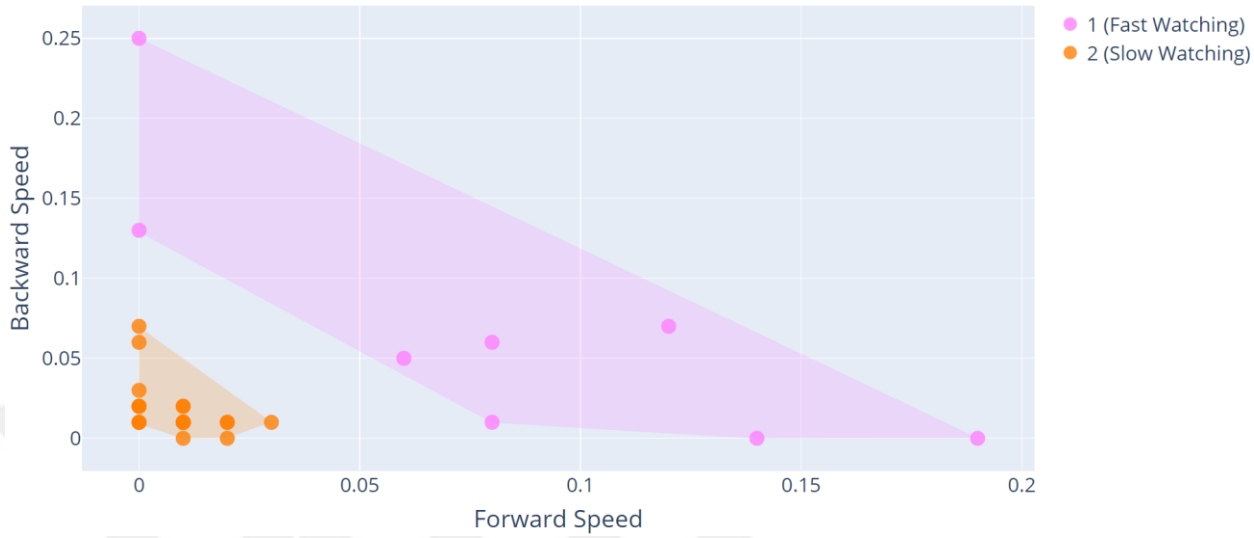



**Figure 12.** Cluster Distribution Based on Forward Speed and Backward Speed

After grouping based on *forward speed* and *backward speed*, the differences among groups could then be compared in terms of learning performance in order to answer the following hypothesis.

*Hypothesis 22:* There is a significant difference in learning performance between groups based on *backward speed* and *forward speed*.

*Null hypothesis:* There is no significant difference in learning performance between groups based on *backward speed* and *forward speed*.

**Table 44.** Robust Linear Mixed Model Results to Predict Learning Performance for Clusters Based on Speed

| ***Response*** | ***Predictor*** | ***Estimate*** | ***SE*** | ***df*** | ***t*** | ***p*** |
|---|---|---|---|---|---|---|
| **Learning performance** | (Intercept) | 75.42 | 3.57 | 31 | 21.12 | .000*** |
| | Cluster (Ref = Fast Watching) Slow Watching | 20.98 | 4.10 | 31 | 5.11 | .000*** |

In order to examine Hypothesis 22, linear mixed model analysis was performed in order to assess the effect of the clusters on the participant students' learning performance. The results show that clusters based on *forward speed* and *backward speed* were able to significantly predict learning performance ($p$ = .000). Results for the Slow Watching cluster produced higher learning performance when compared to the Fast Watching cluster (Estimate = +20.98).

#### 4.2.2.4 Summary Results for Research Question 5

Table 45 presents a summary of all findings related to clustering.

**Table 45.** Summary of Clustering Findings

| *No* | *Hypothesis* | *Clustering Method* | *Result* | | *Accepted* |
|---|---|---|---|---|---|
| | | | *Details* | *p* | |
| 20 | There is a significant difference in learning performance between the groups based on the number of clicks. | two-step cluster analysis | High Clicking behavior results in higher learning performance when compared to Low Clicking (Estimate = +18.23). | .001 | YES |
| 21 | There is a significant difference in learning performance between the groups based on their click preferences. | hierarchical clustering | Backwarder results in higher learning performance when compared to Player (Estimate = +22.03). | .030 | YES |
| 22 | There is a significant difference in learning performance between the groups based on backward speed and forward speed. | two-step cluster analysis | Slow Watching results in higher learning performance when compared to Fast Watching (Estimate = +20.98). | .000 | YES |

### 4.2.3 Predictions Based on Video Viewing Behaviors

*RQ6:* Which video viewing behaviors and to what extent can these behaviors predict students' engagement, satisfaction, and learning performance?

The sixth research question includes five hypotheses (H23, H24, H25, H26, H27) whose content varies according to the respective input (*video clickstream behaviors, behaviors with click reasons*) or output (*satisfaction, engagement, learning performance*). Each is explained in the following related subsections.

*Hypothesis 23:* XGBoost or Robust LMM can significantly predict learning performance better than the baseline model with video clickstream behaviors in terms of three evaluation schemes.

*Null hypothesis:* XGBoost or Robust LMM cannot significantly predict learning performance better than the baseline model with video clickstream behaviors in terms of three evaluation schemes.

Since the aim is to determine whether or not video clickstream behaviors can predict satisfaction, engagement, and learning performance scores alone or together with their respective reasons, two main stages were determined for the predictions. Each stage includes inputs, algorithms, and outputs.

The first uses only video clickstream behaviors based on system data as input, XGBoost (Chen & Guestrin, 2016) and Robust Linear Mixed Model (Koller, 2016) as its algorithms, and satisfaction, engagement, and learning performance as the outputs. The first stage aims to identify whether or not the video clickstream behaviors are sufficient for predicting the variables, and if so, which features are considered as its indicators. The second stage was then followed for those variables which could not be predicted in the first stage, with the aim being to form a prediction together with the respective click reasons. As such, the click reasons were added to the input, thereby making it possible to see if the click reasons are actually necessary in order to achieve a better prediction.

Table 46 presents the main stages for the predictions:

**Table 46.** Main Prediction Stages

| Stage | Inputs | Algorithms | Outputs |
|---|---|---|---|
| 1 | • Video Clickstream Behaviors | • XGBoost<br>• Robust Linear Mixed Model | • Learning performance<br>• Satisfaction<br>• Engagement |
| 2 | • Video Clickstream Behaviors<br>• Click Reasons | • Algorithm which best predicts in Stage 1 | • Outputs not predicted in Stage 1 |

For video clickstream behaviors and click reasons, different features were determined. Table 47 details the features based on video clickstream behaviors. These features were used in the first stage of prediction.

**Table 47.** Features Based on Video Clickstream Behaviors

| No | Feature |
|---|---|
| **1** | Number of Pauses |
| **2** | Number of Forwards |
| **3** | Number of Backwards |
| **4** | Percentage of Pauses |
| **5** | Percentage of Forwards |
| **6** | Percentage of Backwards |
| **7** | Mean Forward Speed |
| **8** | Mean Backward Speed |

Table 48 presents the features based on click reasons. Their scores were determined from the Click Reasons Questionnaire, and their features used in the second stage of prediction together with the features based on video clickstream behaviors.

**Table 48.** Features Based on Click Reasons

| No | Feature |
|---|---|
| **1** | Frequency of Notetaking |
| **2** | Frequency of Losing Concentration |
| **3** | Frequency of Not Understanding Content |
| **4** | Frequency of Facing Language Issues |
| **5** | Frequency of Going Back to Review Important Points |
| **6** | Frequency of Skipping Unimportant Points |
| **7** | Frequency of Wanting to Finish Quickly |

The two stages of prediction are explained in more detail in the following subsections. A summary of all the findings related to prediction are presented in Table 89, which is located at the very end of this section.

#### 4.2.3.1. Learning Performance (Grade)

To predict the learning performance based on XGBoost and Robust LMM, the following main steps were performed respectively:

1. Feature Selection
2. Comparison with Baseline (Random Split)
3. Comparison with Baseline (One Person Leave Out)
4. Comparison with Baseline (One Video Leave Out)

##### 4.2.3.1.1 Learning Performance Prediction Based on Video Clickstream Behaviors (Stage 1)

###### 4.2.3.1.1.1 Feature Selection

Three steps were followed to select the input features for developing the models based on XGBoost, the last two of which were also followed for the Robust LMM algorithm. These steps were (1) parameter tuning (not necessary for Robust LMM), (2) feature ranking, and (3) determining the number of features respectively for both algorithms.

###### 4.2.3.1.1.1.1 Feature Selection for XGBoost

*Step 1: Parameter Tuning for XGBoost*

To find the best parameters, the dataset was split into a training dataset and a testing dataset, using stratified random sampling with five folds (four for training purposes, and one for testing). Therefore, five iterations were performed in total, where one fold was used for testing and four were used for training in each iteration. Stratified mean learning performance (grades/scores) distribution in the original dataset was preserved in the samples. Parameters that gave the best scores on the validation dataset were then selected.

Table 49 presents the parameters and values used for the purpose of hyperparameter tuning the XGBoost algorithm.

**Table 49.** Hyperparameter Tuning (Parameters and Values) for XGBoost

| ***Parameter*** | ***Values*** |
|---|---|
| learning_rate | 0.01, 0.1 |
| max_depth | 3, 5, 7 |
| min_child_weight | 1, 2, 3, 4, 5 |
| subsample | 0.5, 0.7 |
| colsample_bytree | 0.5, 0.7 |
| n_estimators | 400, 500, 600 |
| objective | Mean Squared Error |

*Step 2: Feature Ranking for XGBoost*

For the XGBoost algorithm, the features were each ranked according to the mean feature importance rate over the five iterations.

**Table 50.** Importance of Features with XGBoost for Learning performance

| ***Feature*** | ***Iteration 1*** | ***Iteration 2*** | ***Iteration 3*** | ***Iteration 4*** | ***Iteration 5*** | ***Mean*** | ***Rank*** |
|---|---|---|---|---|---|---|---|
| Number of Pauses | 0.145 | 0.125 | 0.176 | 0.174 | 0.143 | 0.153 | 1 |
| Number of Backwards | 0.134 | 0.108 | 0.133 | 0.181 | 0.195 | 0.150 | 2 |
| Mean Backward Speed | 0.153 | 0.170 | 0.166 | 0.132 | 0.104 | 0.145 | 3 |
| Number of Forwards | 0.144 | 0.086 | 0.121 | 0.138 | 0.131 | 0.124 | 4 |
| Percentage of Pauses | 0.109 | 0.207 | 0.097 | 0.101 | 0.097 | 0.122 | 5 |
| Percentage of Backwards | 0.108 | 0.106 | 0.130 | 0.112 | 0.118 | 0.115 | 6 |
| Mean Forward Speed | 0.142 | 0.102 | 0.119 | 0.083 | 0.109 | 0.111 | 7 |
| Percentage of Forwards | 0.066 | 0.095 | 0.059 | 0.079 | 0.101 | 0.080 | 8 |

*Step 3: Determining the Number of Features for XGBoost*

Root Mean Square Error (RMSE) scores were calculated for each feature set iteratively in order to determine the optimum number of features. The features were added one by one according to their ranking. Since the top three features had the lowest RMSE, and were

also differentiated in terms of their feature importance (see Mean values in Table 54), three features were selected according to their ranks. In addition, Wilcoxon signed-rank test, a non-parametric test for comparing two-related groups (Field, 2018), was performed as a significance test.

Table 51 shows that using the first three features (RMSE = 14.56, MAE = 11.11, *SD* = 9.53) gave better results (less Mean Absolute Error) than when using all eight features (RMSE = 16.38, MAE = 12.75, *SD* = 10.40), *Z* = 605, *p* = .055, even though the results are not significant. This shows that adding more than three features did not improve the model and confirms the feature selection.

**Table 51.** Comparison of Three Features and Eight Features with XGBoost for Learning Performance

| ***No*** | ***Model*** | ***N (observations)*** | ***RMSE*** | ***MAE*** | ***SD*** | ***Z*** | ***p*** |
|---|---|---|---|---|---|---|---|
| 1 | Three-feature model | 42 | 14.56 | 11.11 | 9.53 | 605 | .055 |
| 2 | Eight-feature model | 42 | 16.38 | 12.75 | 10.40 | | |

As a result, the top three features were selected and reordered based on their respective feature importance:

**Table 52.** Importance of Selected Features with XGBoost for Learning Performance

| ***Feature*** | ***Iteration 1*** | ***Iteration 2*** | ***Iteration 3*** | ***Iteration 4*** | ***Iteration 5*** | ***Mean*** | ***Rank*** |
|---|---|---|---|---|---|---|---|
| Number of Pauses | 0.392 | 0.355 | 0.419 | 0.379 | 0.336 | 0.376 | 1 |
| Mean Backward Speed | 0.345 | 0.376 | 0.332 | 0.302 | 0.265 | 0.324 | 2 |
| Number of Backwards | 0.263 | 0.269 | 0.249 | 0.319 | 0.399 | 0.300 | 3 |

* $VIF_1 = 1.53$, $VIF_2 = 1.03$, $VIF_3 = 1.53$

After creating each prediction model, multicollinearity was also checked for each feature with Variance Inflation Factor (VIF). If VIF was less than 5, this reveals there being no multicollinearity (Akinwande et al., 2015).

#### 4.2.3.1.1.1.2 Feature Selection for Robust LMM

*Step 1: Feature Ranking*

For the Robust LMM algorithm, the features were ranked according to their significance value in the model. Table 53 presents the significance of each feature according to its rank.

**Table 53.** Significance of Features with Robust LMM for Learning Performance

| *Rank* | *Variable* | *Estimate* | *SE* | *t* | *p* |
|---|---|---|---|---|---|
| 0 | Intercept | 72.57 | 7.55 | 9.61 | .00 |
| 1 | Mean Backward Speed | -161.51 | 58.99 | -2.74 | .01 |
| 2 | Percentage of Backwards | 37.59 | 18.51 | 2.03 | .05 |
| 3 | Number of Pauses | 0.26 | 0.21 | 1.28 | .21 |
| 4 | Percentage of Pauses | 7.63 | 10.08 | 0.76 | .45 |
| 5 | Percentage of Forwards | 18.99 | 27.72 | 0.69 | .50 |
| 6 | Mean Forward Speed | 0.11 | 0.79 | 0.14 | .66 |
| 7 | Number of Forwards | -53.15 | 119.40 | -0.45 | .89 |
| 8 | Number of Backwards | -0.07 | 0.53 | -0.13 | .90 |

*Step 2: Determining the Number of Features for Robust LMM*

It is possible for one variable to be repressed by another because of interactions between the variables. Therefore, backward feature selection was used in order to select the variables for the final model. Features were then removed from the model according to their rank on a one-by-one basis until the RMSE value increased, or a feature exceeded the *p*-value threshold.

As a result, the first three features were selected since both the RMSE was the lowest and the *p*-values were within threshold for three features. Wilcoxon signed-rank test showed that using three features (RMSE = 15.11, MAE = 11.27, *SD* = 10.19) provided better results (less MAE) than using all eight features (RMSE = 16.40, MAE = 12.66, *SD* = 10.54), $Z = 462$, $p = .185$. Therefore, utilizing all eight features was not needed since the additional features would not improve the model. Therefore, this result also confirmed the feature selection.

**Table 54.** Comparison of Three Features with Eight Features with Robust LMM for Learning Performance

| *No* | *Model* | *N (observations)* | *RMSE* | *MAE* | *SD* | *Z* | *p* |
|---|---|---|---|---|---|---|---|
| 1 | RLMM with three features | 42 | 15.11 | 11.27 | 10.19 | 462 | .185 |
| 2 | RLMM with eight features | 42 | 16.40 | 12.66 | 10.54 | | |

Since the overall goal was to predict the students' learning performance (i.e., midterm exam scores), the *p*-value threshold was selected as .15 rather than .05 as the feature selection stop criteria (Faraway, 2002). Therefore, *number of pauses*, as the third variable in Robust LMM, was also used in the model after backward feature selection (see Table 55). In addition, adding it to the model decreased RMSE. However, adding other variables caused the *p*-value to exceed the .15 threshold, hence the model presented in Table 55 was selected for the Robust LMM. Table 55 presents the fixed effects summary for the final Robust LMM model.

**Table 55.** Fixed Effects of Final Robust LMM for Learning Performance

| *Rank* | *Variable* | *Estimate* | *SE* | *t* | *p* |
|---|---|---|---|---|---|
| 0 | Intercept | 79.83 | 3.78 | 21.14 | .00 |
| 1 | Mean Backward Speed | -162.14 | 51.48 | -3.15 | .00 |
| 2 | Percentage of Backwards | 32.53 | 11.95 | 2.72 | .01 |
| 3 | Number of Pauses | 0.26 | 0.13 | 1.96 | .06 |

* $VIF_1 = 1.09$, $VIF_2 = 1.10$, $VIF_3 = 1.07$

### 4.2.3.1.1.2 Comparing XGBoost and Robust LMM with Baseline (Random Split)

Nested-cross validation was used as the original dataset was split into training and testing datasets, and stratified random sampling was used with five folds for five iterations, which was the same as in Step 1 of Section 4.2.1.1. In each iteration, one fold was set aside as an unseen testing dataset, with the remaining four folds used to validate the model and to find the best parameters for XGBoost and training. The cross-validation parameters and values used for XGBoost are presented in Table 49. For XGBoost regressor, the RMSE score was used to evaluate the model's performance. The results for XGBoost showed that RMSE was 14.76 ($M_{cv}$ = 14.68, $SD_{cv}$ = 1.90).

For Robust LMM, there is no need to perform hypermeter tuning. The results for Robust LMM showed that RMSE was 15.11 ($M_{cv}$ = 14.86, $SD_{cv}$ = 3.73).

However, the RMSE scores for XGBoost and Robust LMM algorithm are not meaningful without the use of a baseline for comparative purposes. Therefore, their respective results were each compared to a baseline model that predicted learning performance in the test dataset using mean learning performance scores from the training dataset.

**Table 56.** Comparison of XGBoost, Robust LMM and Baseline for Learning Performance Prediction (Random Split)

| *No* | *Model* | *N (observations)* | *%RMSE* | *RMSE* | *MAE* | *SD* | $X^2$ | *p* |
|---|---|---|---|---|---|---|---|---|
| 1 | XGBoost | 42 | 17.20 | 14.76 | 11.50 | 9.36 | 12.905 | .002 |
| 2 | Robust LMM | 42 | 17.61 | 15.11 | 11.27 | 10.19 | | |
| 3 | Baseline | 42 | 21.12 | 18.12 | 15.51 | 9.48 | | |

A Friedman test (aka Friedman's ANOVA), a statistical analysis method used for comparing means of several related groups having non-parametric distributions (Field, 2018), was conducted in order to compare the XGBoost and Robust LMM algorithm performances with that of the baseline. The results showed that a significant difference was found to exist ($\chi^2(2) = 12.905$, $p = .002$) (see Table 56). Then, Wilcoxon-Bonferroni post hoc tests were conducted for the purpose of pairwise comparison.

The comparisons showed that both XGBoost (RMSE = 14.76, MAE = 11.50, *SD* = 9.36) and Robust LMM (RMSE = 15.11, MAE = 11.27, *SD* = 10.19) significantly predicted learning performance more accurately than the baseline model (RMSE = 18.12, MAE = 15.51, *SD* = 9.48), *p*(XGBoost-Baseline) = .024, and *p*(RobustLMM-Baseline) = .009. However, there was no significant difference found between XGBoost and Robust LMM, *p*(XGBoost-RobustLMM) = 1.000.

RMSE percentage (aka relative RMSE or normalized RMSE) was also added to the comparison tables (see Table 56, Table 57, and Table 58) (Jamieson et al., 1991; Jeong et al., 2014; Li et al., 2013). The formula for calculating the RMSE percentage is as follows:

$$\%RMSE = \frac{RMSE}{\bar{X}} \text{ x } 100$$

Figure 13 shows the learning performance predictions for each student for each video. Both the XGBoost and Robust LMM models were considered good at predicting high values in a Random Split evaluation scheme. However, the models were shown to predict learning performance of 70 or higher, except for P07, P08, and P15 in Video Lecture 1, and for P02 in Video Lecture 2, which had generally among the lowest scores among the related quizzes. Even if the models predict learning performance with low scores, the

models do not show how low the learning performance scores actually are. Similarly, P02, P07, P10, and also P11 received the lowest scores in the third video lecture. Even if the error rates were considered high for each of them, the models' predictions are generally seen as lower for participants with higher-end scores. This can be explained by the limited prediction range of the models.

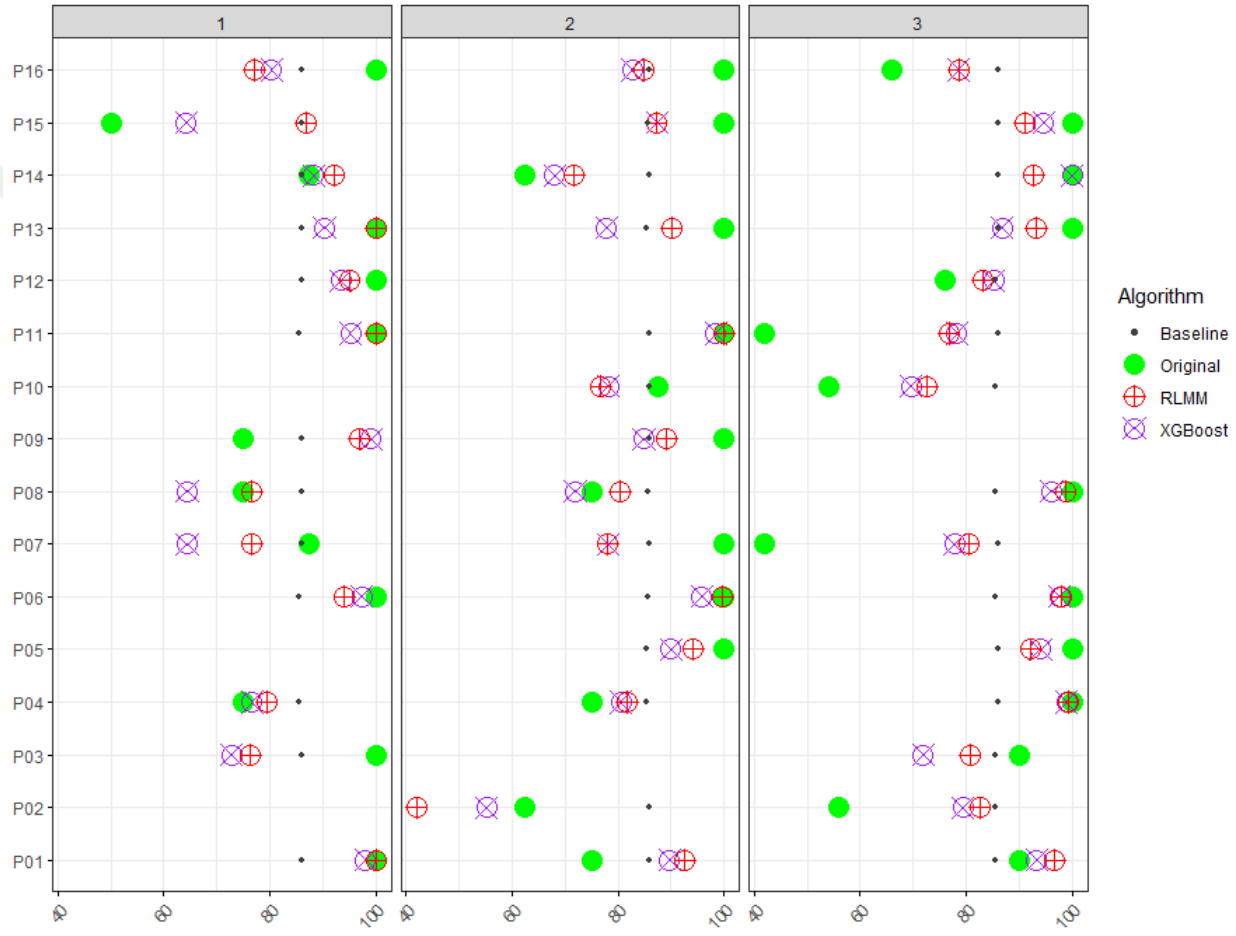



**Figure 13.** Learning Performance Predictions (based on Random Split) for 42 Observations

#### 4.2.3.1.1.3 Comparison of XGBoost and Robust LMM with Baseline (One Person Leave Out)

This evaluation scheme aimed to measure the generalization capability of the proposed models for predicting new students' learning performance using the existing video watching behaviors of other students. One student's watching data for all videos was purposefully left out for testing, with the remainder used for training and validation. This process was performed 16 times in total based on the number of students' data in the dataset, which included a total of 42 observations that were predicted with RMSE.

A baseline model was constructed in order to predict a new student's learning performance for each quiz, based on the mean learning performance of the other students (from the training dataset). Cross-validation parameters and values shown in Table 49 were then used for the XGBoost algorithm. The results for XGBoost showed that RMSE was 15.73 ($M_{cv}$ = 13.60, $SD_{cv}$ = 8.60), whilst for Robust LMM it was 15.26 ($M_{cv}$ = 13.37, $SD_{cv}$ = 7.75).

**Table 57.** Comparison of XGBoost, Robust LMM and Baseline for Learning Performance Prediction (One Person Leave Out)

| *No* | *Model* | *N (observations)* | *%RMSE* | *RMSE* | *MAE* | *SD* | $X^2$ | *p* |
|---|---|---|---|---|---|---|---|---|
| 1 | XGBoost | 42 | 18.33 | 15.73 | 11.63 | 10.71 | 8.143 | .017 |
| 2 | Robust LMM | 42 | 17.79 | 15.26 | 11.52 | 10.14 | | |
| 3 | Baseline | 42 | 22.26 | 19.10 | 16.60 | 9.56 | | |

The Friedman test showed that a significant difference was found to exist between the models ($\chi^2(2) = 8.143$, $p = .017$) (see Table 57). Wilcoxon-Bonferroni post hoc tests showed that both XGBoost (RMSE = 15.73, MAE = 11.63, *SD* = 10.71) and Robust LMM (RMSE = 15.26, MAE = 11.52, *SD* = 10.14) significantly predicted learning performance more accurately than the baseline, according to the "mean" strategy (RMSE = 19.10, MAE = 16.60, *SD* = 9.56), *p*(XGBoost-Baseline) = .045 and *p*(RobustLMM-Baseline) = .000. However, no significant difference was found between XGBoost and Robust LMM, *p*(XGBoost-RobustLMM) = 1.000.

Figure 14 shows the prediction results of the students. The predictions produced for the models showed slightly poorer learning performance scores when compared to the Random Split evaluation scheme. Both models predicted high learning performance scores with a high degree of accuracy; nevertheless, XGBoost was considered to be the more successful in predicting low scores, but less successful in predicting high scores when compared to Robust LMM in this evaluation scheme. Since the number of high scores was more than the number of low scores, Robust LMM was seen to have the better prediction ability. In addition, it is not easy to predict learning performance for students who had zero interaction or only a low number of *backward* click interactions. For example, P02, P03, P07, and also P11 did not perform *backward* click interactions when viewing the third video. As a result, calculating their learning performance was difficult since the number/percentage of *backward clicks* and their mean *backward speed* were all zero.

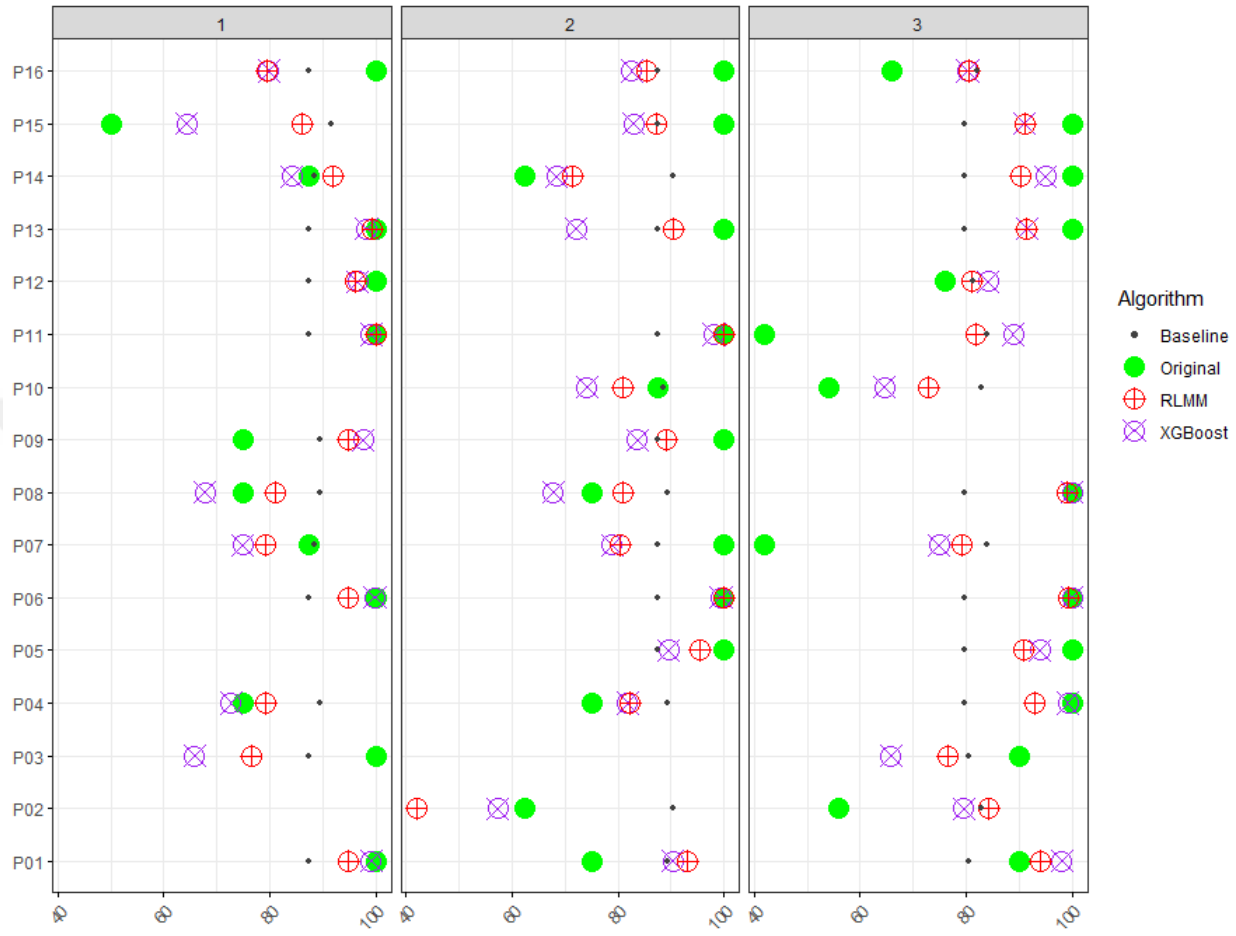



**Figure 14.** Learning Performance Predictions (based on One Person Leave Out) for 42 Observations

#### 4.2.3.1.1.4 Comparing XGBoost and Robust LMM with Baseline (One Video Leave Out)

This evaluation scheme aimed to measure the generalization capability of the models for predicting students' quiz scores after having watched a new video by utilizing their previous video watching behaviors. The students' video watching behaviors for two previous videos were therefore used for the training dataset, with one "new" video used for testing purposes. This process was performed three times, as one for each of the three video lectures, resulting in a total of 42 observations predicted with RMSE. The baseline models were constructed using the mean of the previous quiz scores for each student.

Cross-validation was performed for XGBoost using the parameters and values presented in Table 49. The results for XGBoost showed that RMSE was 17.04 ($M_{cv}$ = 16.57, $SD_{cv}$ = 3.92), whilst for Robust LMM it was 18.24 ($M_{cv}$ = 17.82, $SD_{cv}$ = 4.09). The

comparative results presented in Table 58 show that the XGBoost model outperformed both the Robust LMM and baseline models.

**Table 58.** Comparison of XGBoost, Robust LMM and Baseline for Learning Performance Prediction (One Video Leave Out)

| *No* | *Model* | *N (observations)* | *%RMSE* | *RMSE* | *MAE* | *SD* | $X^2$ | *p* |
|---|---|---|---|---|---|---|---|---|
| 1 | XGBoost | 42 | 19.86 | 17.04 | 11.99 | 12.25 | 11.423 | .003 |
| 2 | Robust LMM | 42 | 21.26 | 18.24 | 13.61 | 12.29 | | |
| 3 | Baseline | 42 | 27.79 | 23.85 | 18.76 | 14.90 | | |

The Friedman test showed that a significant difference was found to exist between the models based on the One Video Leave Out test scenario ($\chi^2(2) = 11.423$, $p = .003$) (see Table 58). Wilcoxon-Bonferroni post hoc tests show that both the XGBoost (RMSE = 17.04, MAE = 11.99, *SD* = 12.25) and Robust LMM (RMSE = 18.24, MAE = 13.61, *SD* = 12.29) models significantly predicted students' learning performance more accurately than the baseline model (RMSE = 23.85, MAE = 18.76, *SD* = 14.90), *p*(XGBoost-Baseline) = .006 and *p*(RobustLMM-Baseline) = .024. However, no significant difference was found between XGBoost and Robust LMM, *p*(XGBoost-RobustLMM) = 1.000.

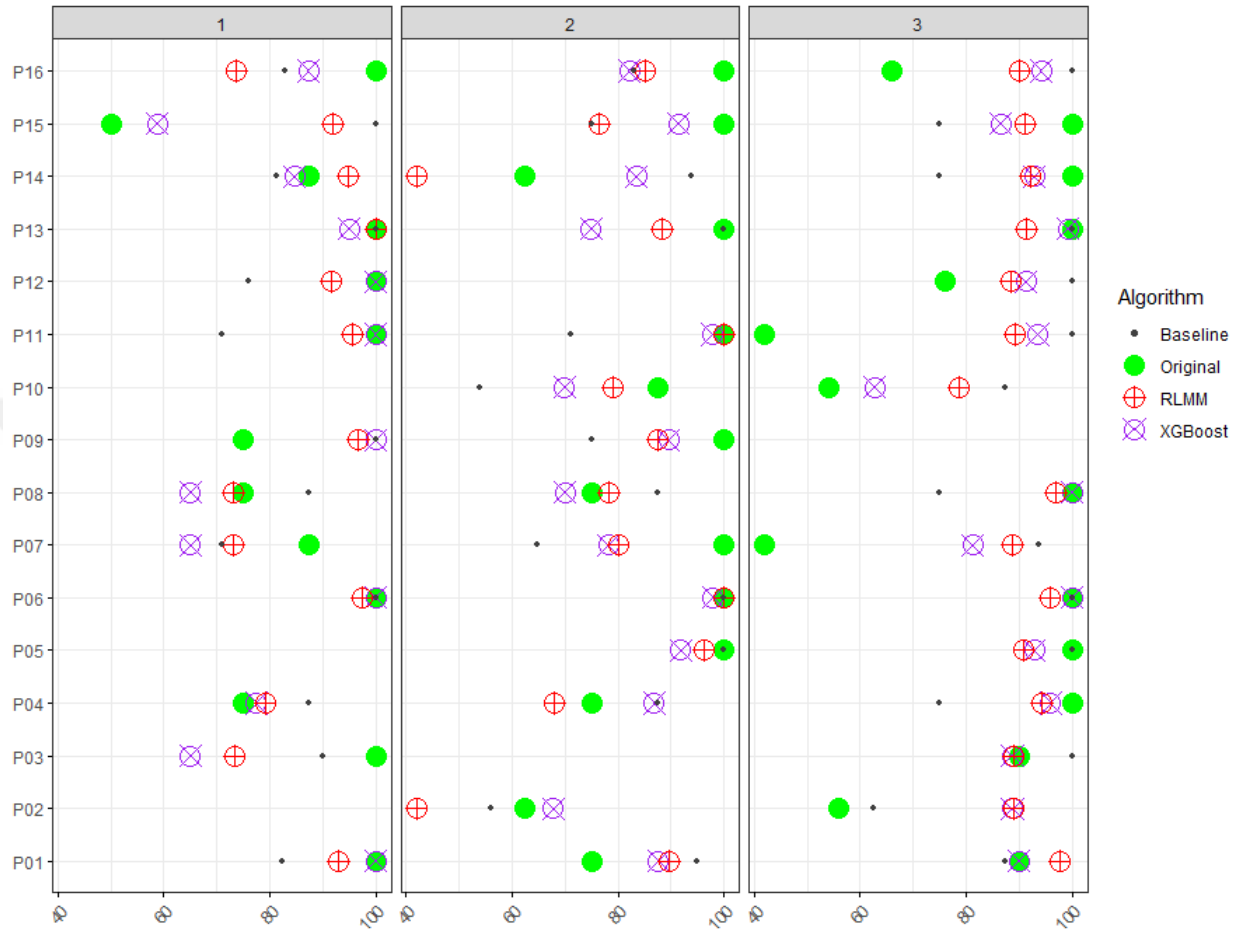



**Figure 15.** Learning Performance Predictions (based on One Video Leave Out) for 42 Observations

Figure 15 shows the individual prediction of the models based on the One Video Leave Out scenario. The success rate in this evaluation scheme was found to be lower than for both the Random Split and One Person Leave Out schemes. XGBoost and Robust LMM are both considered more successful when students receive high scores rather than low scores, as seen in the previous schemes. Correspondingly, it is not easy to predict the students' learning performance based on the One Video Leave Out evaluation scheme when students receive learning performance scores far removed from the general tendency. For example, P07 and P11 both obtained a very low score in the quiz related to the third video, even though both XGBoost and Robust LMM predicted a lower error rate when compared to the baseline model. In addition, the number of student video lectures used within the construction of a model has an important effect on prediction success. Therefore, the success of the models based on the One Video Leave Out scenario would most probably increase if the number of videos was increased.

The results based on the three evaluation validation strategies (Random Split, One Person Leave Out, and One Video Leave Out) showed that both XGBoost and Robust LMM significantly predicted learning performance more accurately than the baseline model, which thereby confirmed Hypothesis 23.

Since learning performance was found to be predicted significantly better than the baseline models in terms of the three evaluation strategies, there was no need to perform the second stage shown in Table 46.

#### 4.2.3.2. Satisfaction

##### 4.2.3.2.1 Satisfaction Prediction Based on Video Clickstream Behaviors (Stage 1)

This subsection tests the following hypothesis:

*Hypothesis 24:* XGBoost or Robust LMM can significantly predict satisfaction better than the baseline model with video clickstream behaviors in terms of three evaluation schemes.

*Null hypothesis:* XGBoost or Robust LMM cannot significantly predict satisfaction better than the baseline model with video clickstream behaviors in terms of three evaluation schemes.

###### 4.2.3.2.1.1 Feature Selection

The same steps were followed for feature selection as for learning performance prediction. While XGBoost includes three steps, Robust LMM includes two steps.

###### 4.2.3.2.1.1.1 Feature Selection for XGBoost

*Step 1: Parameter Tuning for XGBoost*

The same parameters were tuned in Table 49 for the prediction of student satisfaction. However, differently from learning performance prediction, the output from this step was satisfaction.

*Step 2: Feature Ranking with XGBoost*

For satisfaction, the features were ranked according to the mean feature importance rate in the five iterations using the XGBoost algorithm.

**Table 59.** Importance of Features with XGBoost for Satisfaction

| *Feature* | *Iteration 1* | *Iteration 2* | *Iteration 3* | *Iteration 4* | *Iteration 5* | *Mean* | *Rank* |
|---|---|---|---|---|---|---|---|
| Mean Forward Speed | 0.145 | 0.113 | 0.209 | 0.181 | 0.106 | 0.151 | 1 |
| Percentage of Pauses | 0.117 | 0.134 | 0.155 | 0.148 | 0.165 | 0.144 | 2 |
| Number of Pauses | 0.153 | 0.115 | 0.147 | 0.127 | 0.140 | 0.136 | 3 |
| Mean Backward Speed | 0.132 | 0.159 | 0.108 | 0.079 | 0.169 | 0.129 | 4 |
| Number of Backwards | 0.093 | 0.098 | 0.106 | 0.163 | 0.125 | 0.117 | 5 |
| Percentage of Forwards | 0.107 | 0.159 | 0.089 | 0.147 | 0.056 | 0.112 | 6 |
| Number of Forwards | 0.115 | 0.110 | 0.110 | 0.074 | 0.128 | 0.107 | 7 |
| Percentage of Backwards | 0.139 | 0.112 | 0.075 | 0.081 | 0.111 | 0.104 | 8 |

*Step 3: Determining the Number of Features for XGBoost*

The optimum number for feature selection was determined according to RMSE and the model having the optimum number of features was then compared with the model having all features in order to show that using all features was not necessary. First, the features were added to the model one by one in according with their importance, and new RMSE values were calculated after adding each feature so as to find the number of features that had the lowest RMSE value. Then, Wilcoxon signed-rank rank test was performed in order to prove the feature selection.

Table 60 shows that using the first three features (RMSE = 0.76, MAE = 0.64, *SD* = 0.42) produced better results (lower Mean Absolute Error value) than when all eight features were used (RMSE = 0.76, MAE = 0.66, *SD* = 0.44), $Z = 498.500$, $p = .557$, even though the results were not found to be statistically significant. This shows that adding more than three features did not improve the model and thereby confirms the feature selection.

**Table 60.** Comparison of Three Features and Eight Features with XGBoost for Satisfaction

| *No* | *Model* | *N (observations)* | *RMSE* | *MAE* | *SD* | *Z* | *p* |
|---|---|---|---|---|---|---|---|
| 1 | Three-feature model | 42 | .76 | .64 | .42 | 498.500 | .557 |
| 2 | Eight-feature model | 42 | .79 | .66 | .44 | | |

As a result, the top three features were selected and reordered based on feature importance:

**Table 61.** Importance of Selected Features with XGBoost for Satisfaction

| *Feature* | *Iteration 1* | *Iteration 2* | *Iteration 3* | *Iteration 4* | *Iteration 5* | *Mean* | *Rank* |
|---|---|---|---|---|---|---|---|
| Number of Pauses | 0.340 | 0.357 | 0.410 | 0.369 | 0.361 | 0.368 | 1 |
| Percentage of Pauses | 0.321 | 0.316 | 0.336 | 0.306 | 0.369 | 0.330 | 2 |
| Mean Forward Speed | 0.339 | 0.326 | 0.253 | 0.325 | 0.270 | 0.303 | 3 |

* $VIF_1 = 1.08$, $VIF_2 = 1.21$, $VIF_3 = 1.16$

### 4.2.3.2.1.1.2 Feature Selection for Robust LMM

*Step 1: Feature Ranking*

For the Robust LMM algorithm, the features were ranked according to their significance value in the model. Table 62 shows the significance of each feature according to its rank.

**Table 62.** Significance of Features with Robust LMM for Satisfaction (Stage 1)

| *Rank* | *Variable* | *Estimate* | *SE* | *t* | *p* |
|---|---|---|---|---|---|
| 0 | Intercept | 2.93 | 0.40 | 7.33 | .00 |
| 1 | Mean Forward Speed | -11.95 | 5.21 | -2.29 | .03 |
| 2 | Mean Backward Speed | -5.15 | 2.53 | -2.04 | .05 |
| 3 | Number of Backwards | -0.04 | 0.02 | -1.88 | .07 |
| 4 | Number of Pauses | 0.02 | 0.01 | 1.81 | .08 |
| 5 | Number of Forwards | 0.02 | 0.03 | 0.66 | .16 |
| 6 | Percentage of Pauses | -0.72 | 0.50 | -1.43 | .29 |
| 7 | Percentage of Forwards | 0.86 | 1.10 | 0.78 | .44 |
| 8 | Percentage of Backwards | 1.05 | 0.99 | 1.07 | .52 |

*Step 2: Determining the Number of Features for Robust LMM*

**Table 63.** Comparison of Three Features and Eight Features with Robust LMM for Satisfaction (Stage 1)

| *No* | *Model* | *N (observations)* | *RMSE* | *MAE* | *SD* | *Z* | *p* |
|---|---|---|---|---|---|---|---|
| 1 | Three-feature RLMM | 42 | 0.77 | 0.61 | 0.48 | 597.5 | .230 |
| 2 | Eight-feature RLMM | 42 | 0.80 | 0.67 | 0.44 | | |

The Wilcoxon signed-rank test showed that using three features (RMSE = 0.77, MAE = 0.61, *SD* = 0.48) provided better results (lower MAE) than when all eight features were used (RMSE = 0.80, MAE = 0.67, *SD* = 0.44), $Z = 597.5$, $p = .230$. As shown, there was no need to utilize all eight features since adding them did not improve the model. Therefore, this result also confirms the feature selection.

Table 64 presents the fixed effects summary for the final Robust LMM model.

**Table 64.** Fixed Effects of Final Robust LMM for Satisfaction (Stage 1)

| *Rank* | *Variable* | *Estimate* | *SE* | *t* | *p* |
|---|---|---|---|---|---|
| 0 | Intercept | 2.76 | 0.17 | 15.79 | .000 |
| 1 | Mean Forward Speed | -5.46 | 2.71 | -2.01 | .05 |
| 2 | Mean Backward Speed | -4.76 | 2.36 | -2.02 | .05 |

* $VIF_1 = 1.00$, $VIF_2 = 1.00$

### 4.2.3.2.1.2 Comparing XGBoost and Robust LMM with Baseline (Random Split)

The Friedman test showed that no significant difference was found to exist between the models ($\chi^2(2) = 3.854$, $p = .146$) (see Table 65).

**Table 65.** Comparison of XGBoost, Robust LMM and Baseline for Satisfaction Prediction (Random Split in Stage 1)

| *No* | *Model* | *N (observations)* | *%RMSE* | *RMSE* | *MAE* | *SD* | $X^2$ | *p* |
|---|---|---|---|---|---|---|---|---|
| 1 | XGBoost | 42 | 33.03 | 0.83 | 11.50 | 9.36 | 3.854 | .146 |
| 2 | Robust LMM | 42 | 30.57 | 0.77 | 11.27 | 10.19 | | |
| 3 | Baseline | 42 | 31.16 | 0.79 | 15.51 | 9.48 | | |

The prediction details for satisfaction based on the Random Split evaluation scheme are illustrated in Figure 16. As can be seen, XGBoost predicted in a wider range than Robust LMM. However, Robust LMM predicted better than XGBoost. Nevertheless, it produced similar predictions with that of the baseline model.

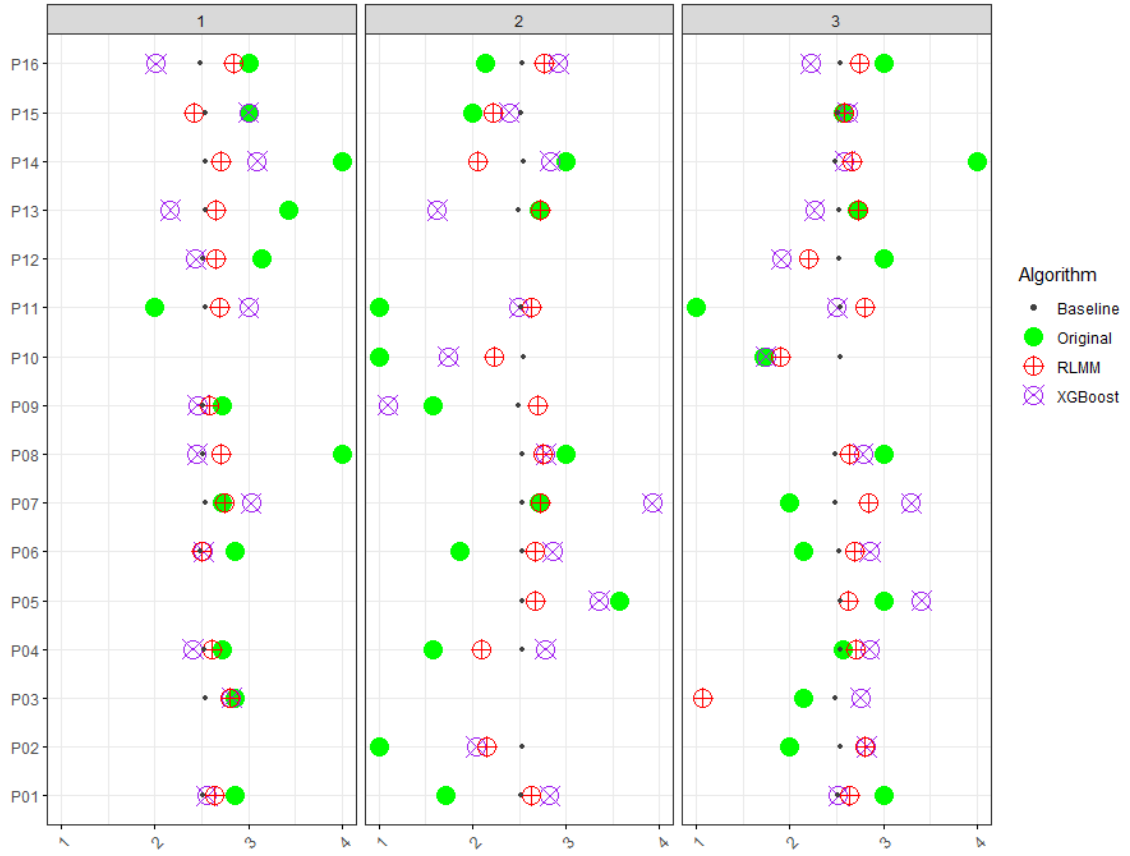



**Figure 16.** Satisfaction Predictions (based on Random Split) for 42 Observations in Stage 1

### 4.2.3.2.1.3 Comparing XGBoost and Robust LMM with Baseline (One Person Leave Out)

The Friedman test showed that no significant difference was found to exist between the models ($\chi^2(2) = 0.190$, $p = .909$) (see Table 66).

**Table 66.** Comparison of XGBoost, Robust LMM and Baseline for Satisfaction Prediction (One Person Leave Out in Stage 1)

| *No* | *Model* | *N (observations)* | *%RMSE* | *RMSE* | *MAE* | *SD* | $X^2$ | *p* |
|---|---|---|---|---|---|---|---|---|
| 1 | XGBoost | 42 | 31.85 | 0.80 | 0.67 | 0.44 | 0.190 | .909 |
| 2 | Robust LMM | 42 | 31.34 | 0.79 | 0.62 | 0.49 | | |
| 3 | Baseline | 42 | 29.13 | 0.74 | 0.58 | 0.45 | | |

Figure 17 shows the predictions of the models for satisfaction based on the One Person Leave Out evaluation scheme. While the RMSE value for Robust LMM increased for this scheme, RMSE with XGBoost was shown to decrease. However, both produced RMSE values that were worse than the baseline model.

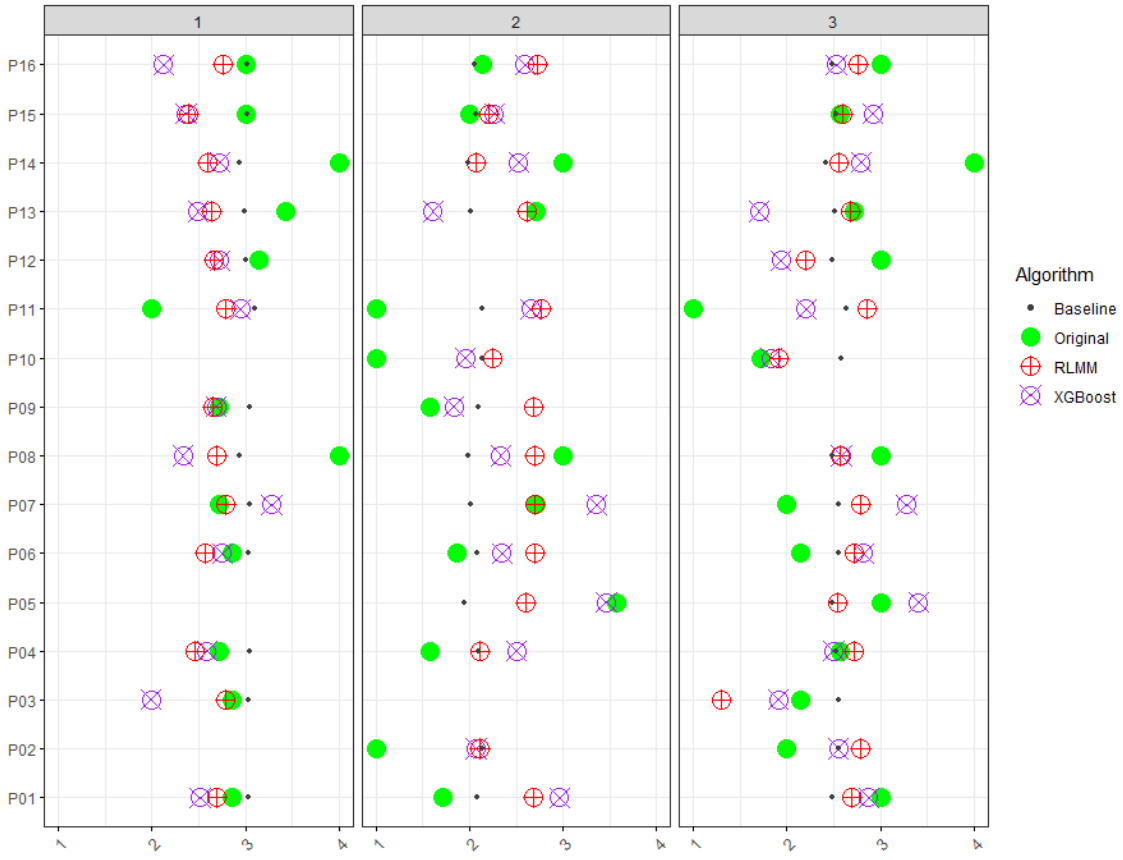



**Figure 17.** Satisfaction Predictions (based on One Person Leave Out) for 42 Observations in Stage 1

#### 4.2.3.2.1.4 Comparing XGBoost and Robust LMM with Baseline (One Video Leave Out)

The Friedman test showed that no significant difference was found to exist between the models ($\chi^2(2) = 2.743$, $p = .254$) (see Table 67).

**Table 67.** Comparison of XGBoost, Robust LMM and Baseline for Satisfaction Prediction (One Video Leave Out in Stage 1)

| *No* | *Model* | *N (observations)* | *%RMSE* | *RMSE* | *MAE* | *SD* | $X^2$ | *p* |
|---|---|---|---|---|---|---|---|---|
| 1 | XGBoost | 42 | 36.04 | 0.91 | 0.79 | 0.52 | 2.743 | .254 |
| 2 | Robust LMM | 42 | 35.06 | 0.89 | 0.72 | 0.52 | | |
| 3 | Baseline | 42 | 27.89 | 0.70 | 0.64 | 0.29 | | |

Figure 18 illustrates predictions of the models for satisfaction for each observation based on the One Video Leave Out evaluation scheme. The results for Robust LMM and XGBoost were found to be the worst among the three evaluation schemes. They predicted large differences compared to the baseline model in terms of RMSE value. Few observations were found to have been correctly predicted, and the predictions were far from the original values.

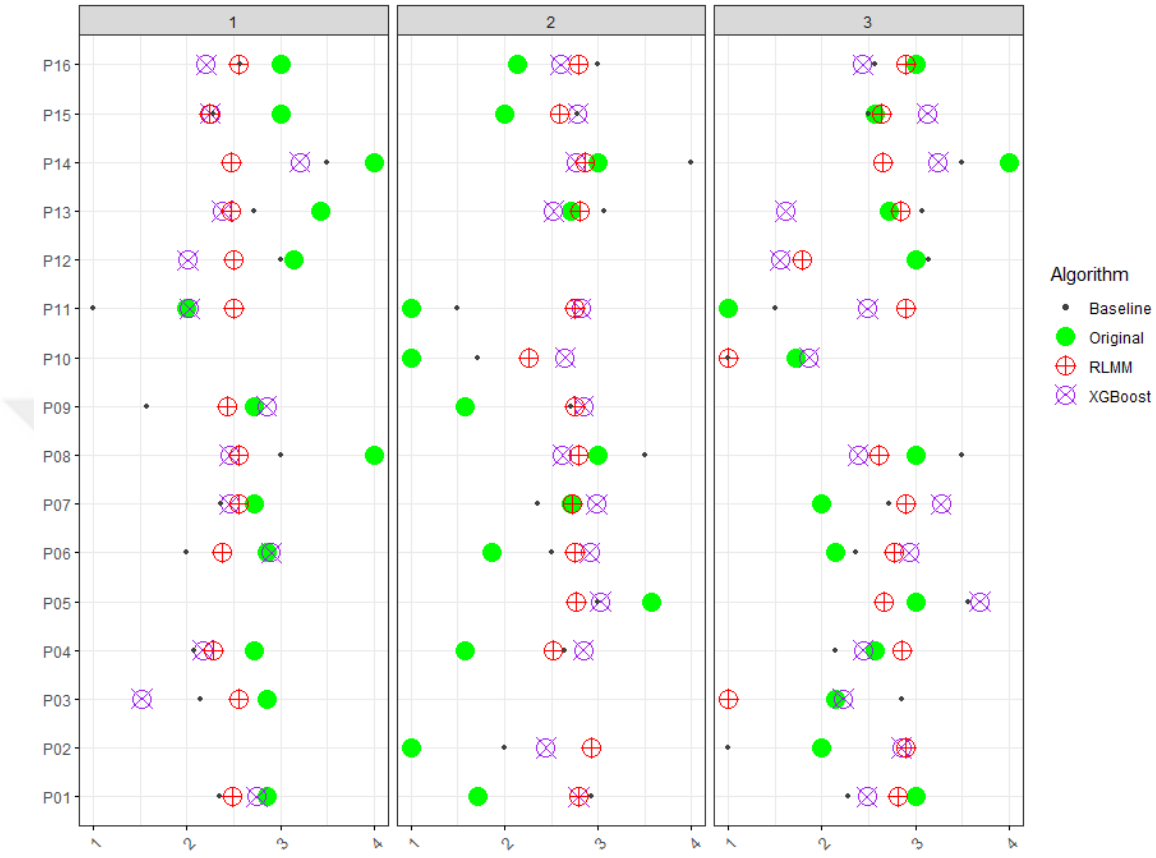



**Figure 18.** Satisfaction Predictions (based on One Video Leave Out) for 42 Observations in Stage 1

The results based on the three evaluation validation strategies (Random Split, One Person Leave Out, and One Video Leave Out) showed that both XGBoost and Robust LMM did not significantly predict satisfaction more accurately than the baseline model, which rejects Hypothesis 24.

#### 4.2.3.2.2. Satisfaction Prediction Based on Video Clickstream Behaviors and Click Reasons (Stage 2)

In the first stage, satisfaction could not be predicted by using only video clickstream behaviors. Therefore, satisfaction was predicted together with the video clickstream behaviors and click reasons in the second stage. As a result, this subsection tests the following hypothesis:

*Hypothesis 25:* The selected algorithm can significantly predict satisfaction better than the baseline model together with video clickstream behaviors and click reasons in terms of three evaluation schemes.

*Null hypothesis:* The selected algorithm cannot significantly predict satisfaction better than the baseline model together with video clickstream behaviors and click reasons in terms of three evaluation schemes.

#### 4.2.3.2.2.1 Algorithm Selection for Stage 2

Robust LMM performed better performance for the prediction of satisfaction in terms of the three evaluation schemes applied in Stage 1:

1. Random Split: RMSE(RLMM) < RMSE(XGBoost)
2. One Person Leave Out: RMSE(RLMM) < RMSE(XGBoost)
3. One Video Leave Out: RMSE(RLMM) < RMSE(XGBoost)

Therefore, Robust LMM was selected for Stage 2.

#### 4.2.3.2.2.2 Feature Selection

*Step 1: Feature Ranking*

For the Robust LMM algorithm, the features were ranked according to their significance value in the model. Table 68 shows the significance of each feature according to its rank.

**Table 68.** Significance of Features with Robust LMM for Satisfaction (Stage 2)

| *Rank* | *Variable* | *Estimate* | *SE* | *t* | *p* |
|---|---|---|---|---|---|
| 0 | Intercept | 4.18 | 0.34 | 12.40 | .00 |
| 1 | Frequency of Wanting to Finish Quickly | -0.39 | 0.08 | -4.98 | .00 |
| 2 | Number of Pauses | 0.02 | 0.01 | 3.28 | .00 |
| 3 | Frequency of Notetaking | -0.20 | 0.07 | -2.86 | .01 |
| 4 | Number of Backwards | -0.04 | 0.01 | -2.56 | .02 |
| 5 | Mean Forward Speed | -9.47 | 3.73 | -2.54 | .02 |
| 6 | Mean Backward Speed | -3.90 | 1.68 | -2.32 | .03 |
| 7 | Frequency of Going Back to Review Important Points | 0.17 | 0.08 | 2.27 | .03 |
| 8 | Frequency of Facing Language Issues | -0.24 | 0.11 | -2.09 | .05 |
| 9 | Percentage of Pauses | -0.50 | 0.32 | -1.55 | .14 |
| 10 | Percentage of Forwards | 1.16 | 0.87 | 1.33 | .20 |
| 11 | Percentage of Backwards | 0.56 | 0.61 | 0.91 | .37 |
| 12 | Frequency of Losing Concentration | 0.08 | 0.13 | 0.65 | .52 |
| 13 | Frequency of Skipping Unimportant Points | 0.04 | 0.10 | 0.44 | .66 |
| 14 | Number of Forwards | 0.00 | 0.02 | -0.09 | .93 |
| 15 | Frequency of Not Understanding Content | 0.01 | 0.13 | 0.08 | .94 |

*Step 2: Determining the Number of Features for Robust LMM*

In the backward feature selection process, nine features were determined. This result was also compared with the model having all features by using Wilcoxon signed-rank test. The result showed that using nine features (RMSE = 0.46, MAE = 0.38, *SD* = 0.26) provided better results (lower MAE) than using all 15 features (RMSE = 0.63, MAE = 0.52, *SD* = 0.37), $Z = 597.5$, $p = .230$. There was no need to utilize all 15 features since adding the additional features did not improve the model. Therefore, this result also confirms the feature selection.

**Table 69.** Comparison of Nine Features and 15 Features with Robust LMM for Satisfaction (Stage 2)

| *No* | *Model* | *N (observations)* | *RMSE* | *MAE* | *SD* | *Z* | *p* |
|---|---|---|---|---|---|---|---|
| 1 | RLMM with nine features | 42 | 0.46 | 0.38 | 0.26 | 659 | .003 |
| 2 | RLMM with 15 features | 42 | 0.63 | 0.52 | 0.37 | | |

Table 70 presents the fixed effects summary of the final Robust LMM model for the prediction of students' satisfaction.

**Table 70.** Fixed Effects of Final Robust LMM for Satisfaction (Stage 2)

| *Rank* | *Variable* | *Estimate* | *SE* | *t* | *p* |
|---|---|---|---|---|---|
| 0 | Intercept | 4.45 | 0.29 | 15.60 | .00 |
| 1 | Frequency of Wanting to Finish Quickly | -0.37 | 0.06 | -5.70 | .00 |
| 2 | Number of Backwards | -0.03 | 0.01 | -3.65 | .00 |
| 3 | Number of Pauses | 0.02 | 0.01 | 2.86 | .01 |
| 4 | Frequency of Facing Language Issues | -0.20 | 0.08 | -2.45 | .02 |
| 5 | Mean Forward Speed | -4.37 | 1.85 | -2.37 | .03 |
| 6 | Frequency of Notetaking | -0.15 | 0.06 | -2.35 | .03 |
| 7 | Frequency of Going Back to Review Important Points | 0.17 | 0.08 | 2.13 | .04 |
| 8 | Mean Backward Speed | -3.13 | 1.56 | -2.00 | .05 |
| 9 | Percentage of Pauses | -0.59 | 0.30 | -2.00 | .06 |

* $VIF_1 = 1.86$, $VIF_2 = 2.20$, $VIF_3 = 2.88$, $VIF_4 = 2.25$, $VIF_5 = 1.34$, $VIF_6 = 2.11$, $VIF_7 = 2.28$, $VIF_8 = 1.08$, $VIF_9 = 1.71$

### 4.2.3.2.2.3 Comparing Robust LMM with Baseline (Random Split)

According to the Random Split evaluation scheme to predict satisfaction, Wilcoxon signed-rank test result showed that Robust LMM (RMSE = 0.46, MAE = 0.38, *SD* = 0.26) significantly predicted students' satisfaction more accurately than the baseline model (RMSE = 0.79, MAE = 0.65, *SD* = 0.46), *Z* = 659, *p* = .009.

**Table 71.** Comparison of Robust LMM and Baseline for Satisfaction Prediction (Random Split in Stage 2)

| *No* | *Model* | *N (observations)* | *%RMSE* | *RMSE* | *MAE* | *SD* | *Z* | *p* |
|---|---|---|---|---|---|---|---|---|
| 1 | Robust LMM | 42 | 18.36 | 0.46 | 0.38 | 0.26 | 659 | .009 |
| 2 | Baseline | 42 | 31.16 | 0.79 | 0.65 | 0.46 | | |

Figure 19 shows the predictions for each observation based on the Random Split evaluation scheme using both video clickstream behaviors, with reasons as the input and students' satisfaction as the output. Robust LMM was shown to perform with better predictions than the baseline. Especially, it predicted values far from the mean that were much better than the baseline.

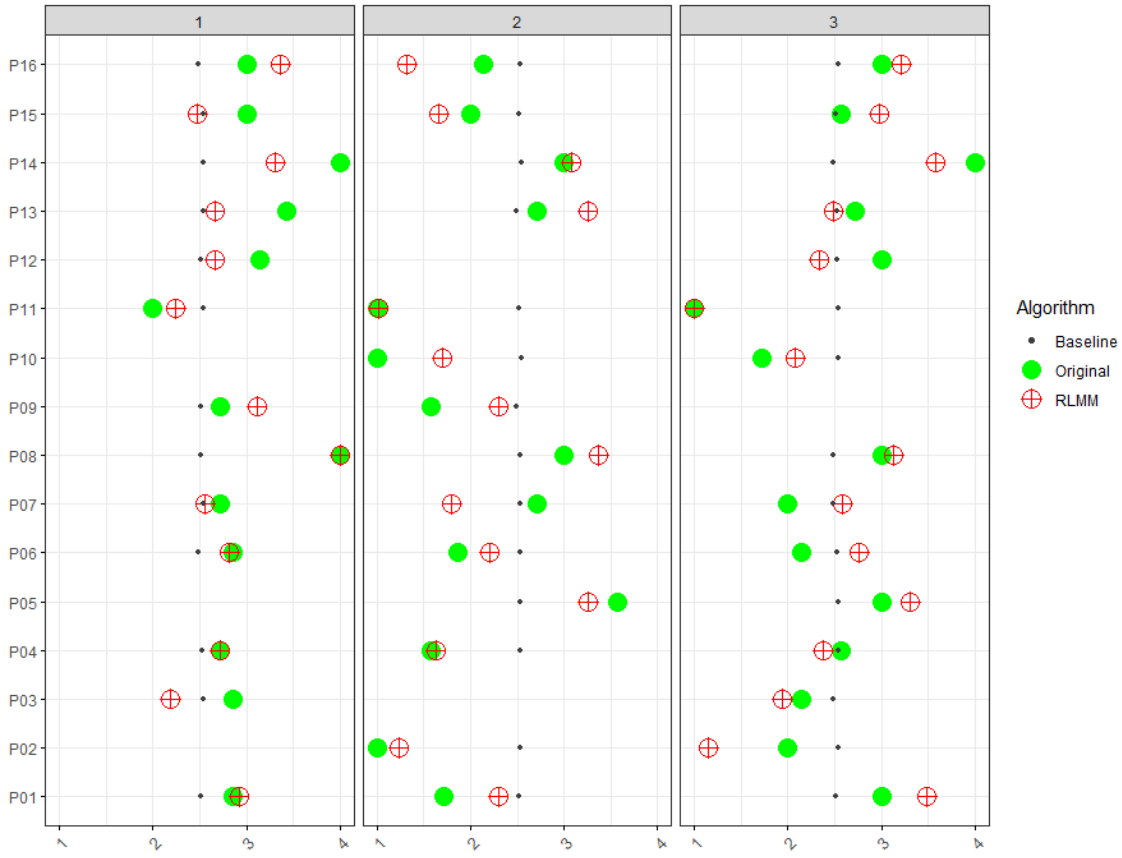



**Figure 19.** Satisfaction Predictions (based on Random Split) for 42 Observations in Stage 2

#### 4.2.3.2.2.4 Comparing Robust LMM with Baseline (One Person Leave Out)

According to One Person Leave Out evaluation scheme used to predict satisfaction, Wilcoxon signed-rank test result showed that Robust LMM (RMSE = 0.45, MAE = 0.39, *SD* = 0.22) significantly predicted students' satisfaction more accurately than the baseline model (RMSE = 0.74, MAE = 0.58, *SD* = 0.45), *Z* = 614, *p* = .042.

**Table 72.** Comparison of Robust LMM and Baseline for Satisfaction Prediction (One Person Leave Out in Stage 2)

| *No* | *Model* | *N (observations)* | *%RMSE* | *RMSE* | *MAE* | *SD* | *Z* | *p* |
|---|---|---|---|---|---|---|---|---|
| 1 | Robust LMM | 42 | 17.66 | 0.45 | 0.39 | 0.22 | 614 | .042 |
| 2 | Baseline | 42 | 29.13 | 0.74 | 0.58 | 0.45 | | |

Figure 20 shows the predictions for each observation based on the One Person Leave Out scheme, with both video clickstream behaviors and their reasons used to predict the students' satisfaction. The predictions of Robust LMM in this scheme were found to be slightly better than those of the Random Split scheme. Robust LMM predicted the values of low satisfaction as low and the values of high satisfaction as high, even where it could not predict the exact value.

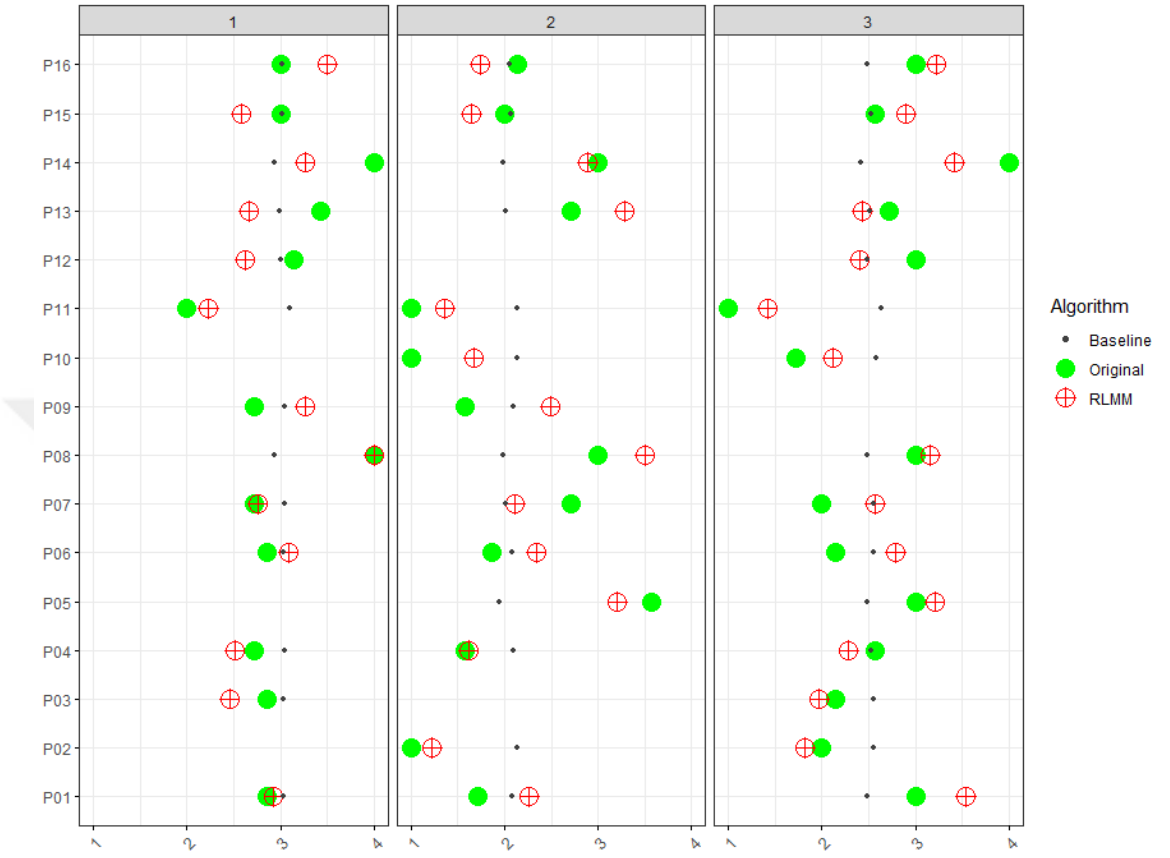



**Figure 20.** Satisfaction Predictions (based on One Person Leave Out) for 42 Observations in Stage 2

### 4.2.3.2.2.5 Comparing Robust LMM with Baseline (One Video Leave Out)

According to the One Video Leave Out evaluation scheme used to predict satisfaction, Wilcoxon signed-rank test result showed that Robust LMM (RMSE = 0.43, MAE = 0.37, *SD* = 0.22) significantly predicted students' satisfaction more accurately than the baseline model (RMSE = 0.70, MAE = 0.64, *SD* = 0.29), *Z* = 739, *p* = .000.

**Table 73.** Comparison of Robust LMM and Baseline for Satisfaction Prediction (One Video Leave Out in Stage 2)

| *No* | *Model* | *N (observations)* | *%RMSE* | *RMSE* | *MAE* | *SD* | *Z* | *p* |
|---|---|---|---|---|---|---|---|---|
| 1 | Robust LMM | 42 | 17.07 | 0.43 | 0.37 | 0.22 | 739 | .000 |
| 2 | Baseline | 42 | 27.89 | 0.70 | 0.64 | 0.29 | | |

Figure 21 shows the predictions for each observation based on the One Video Leave Out scheme, with both video clickstream behaviors and their reasons used to predict the

students' satisfaction. The predictions of Robust LMM in this scheme were found to be the best among the three evaluation schemes. Robust LMM predicted the lowest satisfaction in this scheme better than the other schemes. For example, the predictions for P11 in Video Lecture 1, P10 and P11 in Video Lecture 2, and P02, P10, and P11 in Video Lecture 3 were all better when compared to the predictions in the One Person Leave Out scheme.

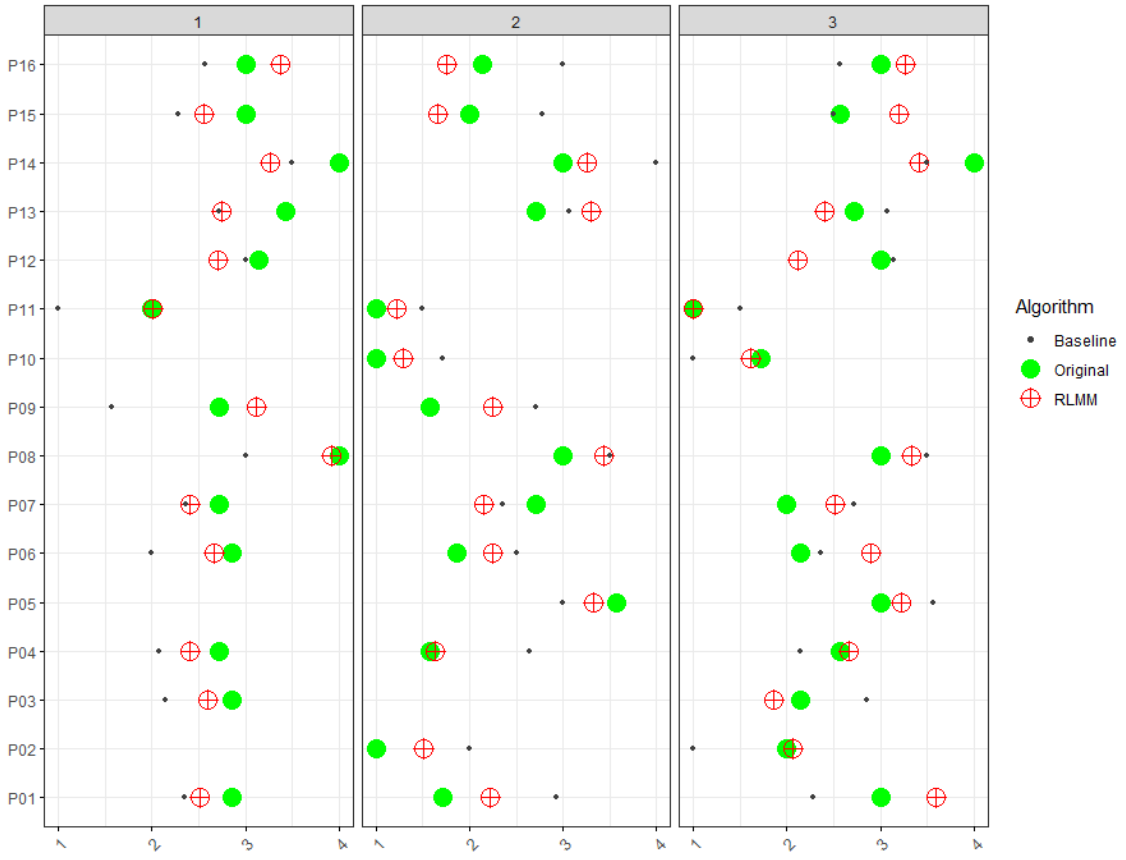



**Figure 21.** Satisfaction Predictions (based on One Video Leave Out) for 42 Observations in Stage 2

The results based on the three evaluation validation strategies (Random Split, One Person Leave Out, and One Video Leave Out) showed that by using both video clickstream behaviors and their reasons, the selected algorithm significantly predicted satisfaction more accurately than the baseline model, which confirms Hypothesis 25.

#### 4.2.3.3. Engagement

##### 4.2.3.3.1 Engagement Prediction Based on Video Clickstream Behaviors (Stage 1)

This subsection tests the following hypothesis:

*Hypothesis 26:* XGBoost or Robust LMM can significantly predict engagement better than the baseline model with video clickstream behaviors in terms of three evaluation schemes.

*Null hypothesis:* XGBoost or Robust LMM cannot significantly predict engagement better than the baseline model with video clickstream behaviors in terms of three evaluation schemes.

###### 4.2.3.3.1.1 Feature Selection

The same steps were followed as for learning performance prediction in the feature selection for predicting engagement. While XGBoost includes three steps, Robust LMM includes two steps.

###### 4.2.3.3.1.1.1 Feature Selection for XGBoost

*Step 1: Parameter Tuning for XGBoost*

The same parameters were used as shown in Table 49 for the prediction of engagement.

*Step 2: Feature Ranking with XGBoost*

For engagement, the features were ranked according to their mean feature importance rate in the five iterations using the XGBoost algorithm.

**Table 74.** Importance of Features with XGBoost for Engagement

| *Feature* | *Iteration 1* | *Iteration 2* | *Iteration 3* | *Iteration 4* | *Iteration 5* | *Mean* | *Rank* |
|---|---|---|---|---|---|---|---|
| Number of Pauses | 0.185 | 0.099 | 0.112 | 0.163 | 0.156 | 0.143 | 1 |
| Mean Forward Speed | 0.135 | 0.189 | 0.118 | 0.137 | 0.108 | 0.137 | 2 |
| Percentage of Pauses | 0.132 | 0.114 | 0.133 | 0.156 | 0.106 | 0.128 | 3 |
| Number of Backwards | 0.119 | 0.095 | 0.127 | 0.135 | 0.144 | 0.124 | 4 |
| Number of Forwards | 0.102 | 0.129 | 0.127 | 0.134 | 0.106 | 0.120 | 5 |
| Mean Backward Speed | 0.102 | 0.131 | 0.139 | 0.085 | 0.139 | 0.119 | 6 |
| Percentage of Backwards | 0.098 | 0.121 | 0.112 | 0.104 | 0.139 | 0.115 | 7 |
| Percentage of Forwards | 0.126 | 0.121 | 0.132 | 0.086 | 0.100 | 0.113 | 8 |

*Step 3: Determining the Number of Features for XGBoost*

The RMSE score was lowest when using the first two features to predict engagement with XGBoost. Therefore, the model having two features was then compared with the model that had all eight features by using Wilcoxon signed-rank test.

Table 75 shows that using the first two features (RMSE = 1.292, MAE = 1.04, *SD* =0.77) gave better results (lower MAE) than when using all eight features (RMSE = 1.295, MAE = 1.02, *SD* = 0.81), $Z = 417.5$, $p = .671$, even though the results were not statistically significant. This shows that adding more than three features did not improve the model, which thereby confirms the feature selection.

**Table 75.** Comparison of Two Features and Eight Features with XGBoost for Engagement

| *No* | *Model* | *N (observations)* | *RMSE* | *MAE* | *SD* | *Z* | *p* |
|---|---|---|---|---|---|---|---|
| 1 | Two-feature model | 42 | 1.292 | 1.04 | 0.77 | 417.5 | .671 |
| 2 | Eight-feature model | 42 | 1.295 | 1.02 | 0.81 | | |

As a result, the top two features were selected and reordered based on their feature importance:

**Table 76.** Importance of Selected Features with XGBoost for Engagement

| *Feature* | *Iteration 1* | *Iteration 2* | *Iteration 3* | *Iteration 4* | *Iteration 5* | *Mean* | *Rank* |
|---|---|---|---|---|---|---|---|
| Number of Pauses | 0.522 | 0.593 | 0.582 | 0.624 | 0.504 | 0.565 | 1 |
| Mean Forward Speed | 0.478 | 0.407 | 0.418 | 0.376 | 0.496 | 0.435 | 2 |

* $VIF_1 = 1.03$, $VIF_2 = 1.03$

### 4.2.3.3.1.1.2 Feature Selection for Robust LMM

*Step 1: Feature Ranking*

For the Robust LMM algorithm, the features were ranked according to their significance value in the model. Table 77 shows the significance of each feature according to its rank. The aim of feature ranking for Robust LMM was to reveal the importance of all the features as part of the bigger picture. During the backward feature selection, the ranks of features may change among the remaining features; hence, why the final model did not include the first two variables (see Table 77) as they were eliminated during the backward feature selection process.

**Table 77.** Significance of Features with Robust LMM for Engagement (Stage 1)

| *Rank* | *Variable* | *Estimate* | *SE* | *t* | *p* |
|---|---|---|---|---|---|
| 0 | Intercept | 3.71 | 0.62 | 5.96 | .00 |
| 1 | Number of Backwards | -0.08 | 0.04 | -1.87 | .07 |
| 2 | Percentage of Backwards | 2.72 | 1.53 | 1.78 | .09 |
| 3 | Mean Forward Speed | -12.88 | 9.85 | -1.31 | .20 |
| 4 | Mean Backward Speed | -5.75 | 4.87 | -1.18 | .25 |
| 5 | Percentage of Pauses | -0.70 | 0.83 | -0.84 | .41 |
| 6 | Number of Pauses | 0.01 | 0.02 | 0.82 | .42 |
| 7 | Number of Forwards | 0.03 | 0.06 | 0.44 | .67 |
| 8 | Percentage of Forwards | 0.87 | 2.29 | 0.38 | .71 |

*Step 2: Determining the Number of Features for Robust LMM*

In the backward feature selection process, one feature was selected for predicting engagement with video clickstream behaviors. This result was also compared with the model having all eight features by using Wilcoxon signed-rank test. The result showed that using one feature (RMSE = 1.22, MAE = 0.95, *SD* = 0.78) provided significantly better results (less MAE) than when using all eight features (RMSE = 1.48, MAE = 1.19, *SD* = 0.90), *Z* = 707, *p* = 001. This showed that using all eight features resulted in a worse RMSE value, which thereby confirms the feature selection.

**Table 78.** Comparison of One Feature and Eight Features with Robust LMM for Engagement (Stage 1)

| *No* | *Model* | *N (observations)* | *RMSE* | *MAE* | *SD* | *Z* | *p* |
|---|---|---|---|---|---|---|---|
| 1 | RLMM with one feature | 42 | 1.22 | 0.95 | 0.78 | 707 | .001 |
| 2 | RLMM with eight features | 42 | 1.48 | 1.19 | 0.90 | | |

Table 79 presents the fixed effects summary of the final Robust LMM model for the prediction of student engagement.

**Table 79.** Fixed Effects of Final Robust LMM for Engagement (Stage 1)

| *Rank* | *Variable* | *Estimate* | *SE* | *t* | *p* |
|---|---|---|---|---|---|
| 0 | Intercept | 3.51 | 0.22 | 16.24 | .000 |
| 1 | Mean Forward Speed | -6.97 | 4.65 | -2.01 | -1.50 |

### 4.2.3.3.1.2 Comparing XGBoost and Robust LMM with Baseline (Random Split)

According to the Random Split evaluation scheme used to predict engagement with video clickstream behaviors, the Friedman test showed that a significant difference was found to exist between the models ($\chi^2(2) = 14.619$, $p = .001$) (see Table 80). Then, Wilcoxon-Bonferroni post hoc tests were conducted for the purpose of pairwise comparison.

The comparisons showed that Robust LMM (RMSE = 1.22, MAE = 0.95, *SD* = 0.78) significantly predicted engagement more accurately than XGBoost (RMSE = 1.29, MAE = 1.05, *SD* = 0.77), *p*(XGBoost-RobustLMM) = .03. However, neither Robust LMM nor XGBoost significantly predicted engagement more accurately than the baseline model (RMSE = 1.22, MAE = 0.97, *SD* = 0.76), *p*(XGBoost-Baseline) = .354 and *p*(RobustLMM-Baseline) = 1.000.

**Table 80.** Comparison of XGBoost, Robust LMM and Baseline for Engagement Prediction (Random Split in Stage 1)

| *No* | *Model* | *N (observations)* | *%RMSE* | *RMSE* | *MAE* | *SD* | $X^2$ | *p* |
|---|---|---|---|---|---|---|---|---|
| 1 | XGBoost | 42 | 37.92 | 1.29 | 1.05 | 0.77 | 14.619 | .001 |
| 2 | Robust LMM | 42 | 35.73 | 1.22 | 0.95 | 0.78 | | |
| 3 | Baseline | 42 | 35.81 | 1.22 | 0.97 | 0.76 | | |

Figure 22 illustrates predictions for each observation based on the Random Split evaluation scheme, using only video clickstream behaviors as input and student engagement as its output. As can be seen, Robust LMM performed better predictions than XGBoost, while Robust LMM made almost the same predictions as the baseline model. However, the predictions including the baseline model were far away from the original values.




**Figure 22.** Engagement Predictions (based on Random Split) for 42 Observations in Stage 1

#### 4.2.3.3.1.3 Comparing XGBoost and Robust LMM with Baseline (One Person Leave Out)

According to One Person Leave Out evaluation scheme used to predict engagement with video clickstream behaviors, the Friedman test showed that a significant difference was found to exist between the models ($\chi^2(2) = 8.333$, $p = .016$) (see Table 81). Then, Wilcoxon-Bonferroni post hoc tests were conducted for the purpose of pairwise comparison.

The comparisons showed that Robust LMM (RMSE = 1.27, MAE = 1.01, *SD* = 0.78) significantly predicted engagement more accurately than XGBoost (RMSE = 1.41, MAE = 1.15, *SD* = 0.82), *p*(XGBoost-RobustLMM) = .01. However, neither Robust LMM nor XGBoost significantly predicted engagement more accurately than the baseline model (RMSE = 1.12, MAE = 0.91, *SD* = 0.67), *p*(XGBoost-Baseline) = .06 and *p*(RobustLMM-Baseline) = 1.000

**Table 81.** Comparison of XGBoost, Robust LMM, and Baseline for Engagement Prediction (One Person Leave Out in Stage 1)

| *No* | *Model* | *N (observations)* | *%RMSE* | *RMSE* | *MAE* | *SD* | $X^2$ | *p* |
|---|---|---|---|---|---|---|---|---|
| 1 | XGBoost | 42 | 41.31 | 1.41 | 1.15 | 0.82 | 8.333 | .016 |
| 2 | Robust LMM | 42 | 37.26 | 1.27 | 1.01 | 0.78 | | |
| 3 | Baseline | 42 | 32.88 | 1.12 | 0.91 | 0.67 | | |

Figure 23 shows the predictions of the models for engagement for each observation based on the One Person Leave Out evaluation scheme. While RMSE for both Robust LMM and XGBoost increased in this scheme, RMSE decreased for the baseline model. Both predicted with lower accuracy than the baseline model, especially in the first and second video lectures.

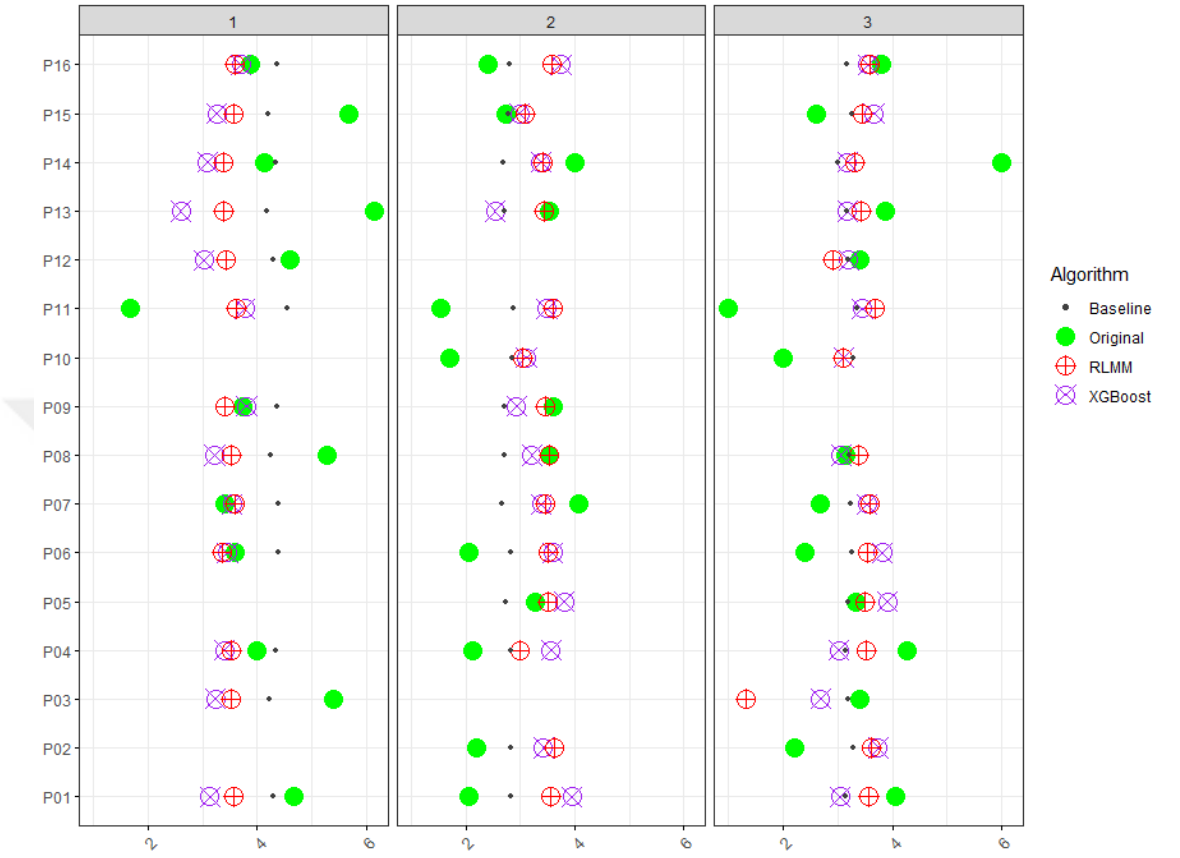



**Figure 23.** Engagement Predictions (based on One Person Leave Out) for 42 Observations in Stage 1

#### 4.2.3.3.1.4 Comparing XGBoost and Robust LMM with Baseline (One Video Leave Out)

According to the One Person Leave Out evaluation scheme used to predict engagement with video clickstream behaviors, the Friedman test showed that no significant difference was found to exist between the models ($\chi^2(2) = 1.000$, $p = .607$) (see Table 82).

**Table 82.** Comparison of XGBoost, Robust LMM and Baseline for Engagement Prediction (One Video Leave Out in Stage 1)

| *No* | *Model* | *N (observations)* | *%RMSE* | *RMSE* | *MAE* | *SD* | $X^2$ | *p* |
|---|---|---|---|---|---|---|---|---|
| 1 | XGBoost | 42 | 41.76 | 1.42 | 1.15 | 0.85 | 1.000 | .607 |
| 2 | Robust LMM | 42 | 41.69 | 1.42 | 1.16 | 0.84 | | |
| 3 | Baseline | 42 | 37.40 | 1.28 | 1.04 | 0.75 | | |

Figure 24 shows the predictions of the models for engagement for each observation based on the One Video Leave Out evaluation scheme. The results for the models including the baseline model were the worst seen among the three evaluation schemes. The estimates differentiated greatly from the actual values for all of the models.

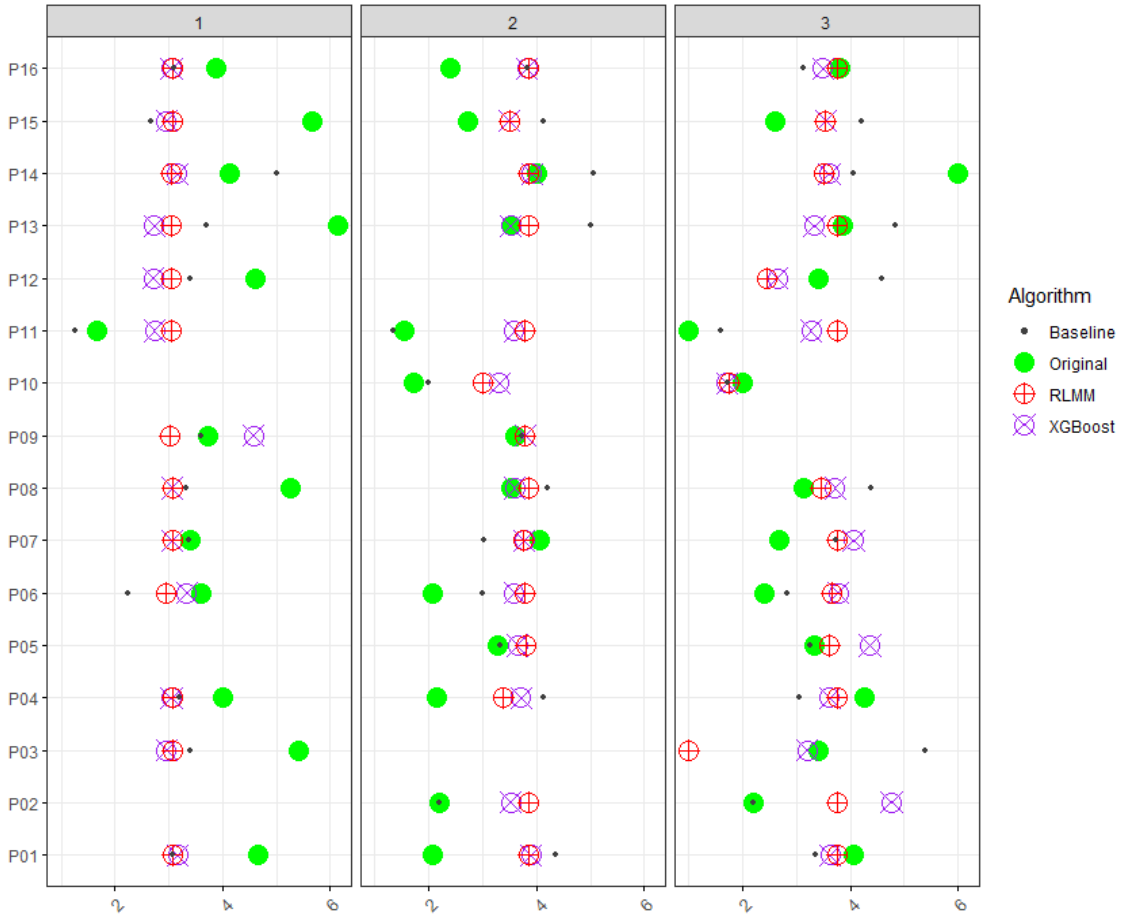



**Figure 24.** Engagement Predictions (based on One Video Leave Out) for 42 Observations in Stage 1

The results based on the three evaluation validation strategies (Random Split, One Person Leave Out, and One Video Leave Out) showed that both XGBoost and Robust LMM did not significantly predict engagement with video clickstream behaviors more accurately than the baseline model, which thereby rejects Hypothesis 26.

#### 4.2.3.3.2 Engagement Prediction Based on Video Clickstream Behaviors and Click Reasons (Stage 2)

The first stage revealed that it was not possible to predict engagement using only video clickstream behaviors. Thus, click reasons were also added as input to the second stage. As a result, this subsection tests the following hypothesis:

*Hypothesis 27:* The selected algorithm can significantly predict engagement better than the baseline model together with video clickstream behaviors and click reasons in terms of three evaluation schemes.

*Null hypothesis:* The selected algorithm cannot significantly predict engagement better than the baseline model together with video clickstream behaviors and click reasons in terms of three evaluation schemes.

#### 4.2.3.3.2.1 Algorithm Selection for Stage 2

Robust LMM performed better for the prediction of engagement in terms of the three evaluation schemes in Stage 1:

1. Random Split: RMSE(RLMM) < RMSE(XGBoost)
2. One Person Leave Out: RMSE(RLMM) < RMSE(XGBoost)
3. One Video Leave Out: RMSE(RLMM) < RMSE(XGBoost)

Therefore, Robust LMM was selected for Stage 2.

#### 4.2.3.3.2.2 Feature Selection

*Step 1: Feature Ranking*

For the Robust LMM algorithm, the features were ranked according to their significance value in the model. Table 83 shows the significance of each feature according to its rank.

**Table 83.** Significance of Features with Robust LMM for Engagement (Stage 2)

| *Rank* | *Variable* | *Estimate* | *SE* | *t* | *p* |
|---|---|---|---|---|---|
| 0 | Intercept | 5.29 | 0.79 | 6.71 | .00 |
| 1 | Frequency of Notetaking | -0.29 | 0.17 | -1.77 | .09 |
| 2 | Number of Backwards | -0.06 | 0.04 | -1.61 | .12 |
| 3 | Frequency of Going Back to Review Important Points | 0.27 | 0.19 | 1.41 | .17 |
| 4 | Frequency of Wanting to Finish Quickly | -0.24 | 0.19 | -1.31 | .21 |
| 5 | Number of Pauses | 0.02 | 0.02 | 1.18 | .25 |
| 6 | Mean Forward Speed | -10.58 | 9.19 | -1.15 | .26 |
| 7 | Frequency of Skipping Unimportant Points | 0.21 | 0.25 | 0.85 | .40 |
| 8 | Frequency of Losing Concentration | -0.24 | 0.31 | -0.78 | .44 |
| 9 | Percentage of Backwards | 0.99 | 1.42 | 0.70 | .50 |
| 10 | Frequency of Not Understanding Content | -0.19 | 0.31 | -0.61 | .55 |
| 11 | Percentage of Pauses | -0.37 | 0.74 | -0.50 | .62 |
| 12 | Number of Forwards | 0.02 | 0.06 | 0.38 | .71 |
| 13 | Frequency of Facing Language Issues | -0.10 | 0.28 | -0.37 | .71 |
| 14 | Mean Backward Speed | -1.36 | 4.26 | -0.32 | .75 |
| 15 | Percentage of Forwards | 0.45 | 2.24 | 0.20 | .84 |

*Step 2: Determining the Number of Features for Robust LMM*

In the backward feature selection process, two features were determined for the predicting of engagement with both video clickstream behaviors and their reasons. This result was also compared with the model having all 15 features by using Wilcoxon signed-rank test.

The result shows that using two features (RMSE = 1.03, MAE = 0.81, *SD* = 0.65) provided significantly better results (less MAE) than using all 15 features (RMSE = 1.51, MAE = 1.20, *SD* = 0.92), *Z* = 659, *p* = .009. Using all 15 features made the model worse; hence, this result also confirms the feature selection.

**Table 84.** Comparison of Two Features and 15 features with Robust LMM for Engagement (Stage 2)

| *No* | *Model* | *N (observations)* | *RMSE* | *MAE* | *SD* | *Z* | *p* |
|---|---|---|---|---|---|---|---|
| 1 | RLMM with two features | 42 | 1.03 | 0.81 | 0.65 | 659 | .009 |
| 2 | RLMM with 15 features | 42 | 1.51 | 1.20 | 0.92 | | |

Table 85 presents the fixed effects summary of the final Robust LMM model for the prediction of student engagement.

**Table 85.** Fixed Effects of Final Robust LMM for Engagement (Stage 2)

| *Rank* | *Variable* | *Estimate* | *SE* | *t* | *p* |
|---|---|---|---|---|---|
| 0 | Intercept | 5.17 | 0.43 | 11.98 | .00 |
| 1 | Frequency of Losing Concentration | -0.62 | 0.14 | -4.45 | .00 |
| 2 | Mean Forward Speed | -8.04 | 3.43 | -2.34 | .02 |

* $VIF_1 = 1.00$, $VIF_2 = 1.00$

### 4.2.3.3.2.3 Comparing Robust LMM with Baseline (Random Split)

According to the Random Split evaluation scheme used to predict engagement with both video clickstream behaviors and their reasons, Wilcoxon signed-rank test result showed that Robust LMM (RMSE = 1.03, MAE = 0.81, *SD* = 0.65) did not significantly predict student engagement more accurately than the baseline model (RMSE = 1.22, MAE = 0.97, *SD* = 0.76), *Z* = 603, *p* = .058.

**Table 86.** Comparison of XGBoost, Robust LMM and Baseline for Engagement Prediction (Random Split in Stage 2)

| *No* | *Model* | *N (observations)* | *%RMSE* | *RMSE* | *MAE* | *SD* | *Z* | *p* |
|---|---|---|---|---|---|---|---|---|
| 1 | Robust LMM | 42 | 30.26 | 1.03 | 0.81 | 0.65 | 603 | .058 |
| 2 | Baseline | 42 | 35.81 | 1.22 | 0.97 | 0.76 | | |

Figure 25 shows the predictions for each observation based on the Random Split scheme with both video clickstream behaviors and their reasons to predict student engagement. The predictions of Robust LMM in this scheme were found to be considerably better than those for the Random Split scheme with only video clickstream behaviors. However, no significant difference was found between Robust LMM and the baseline model.

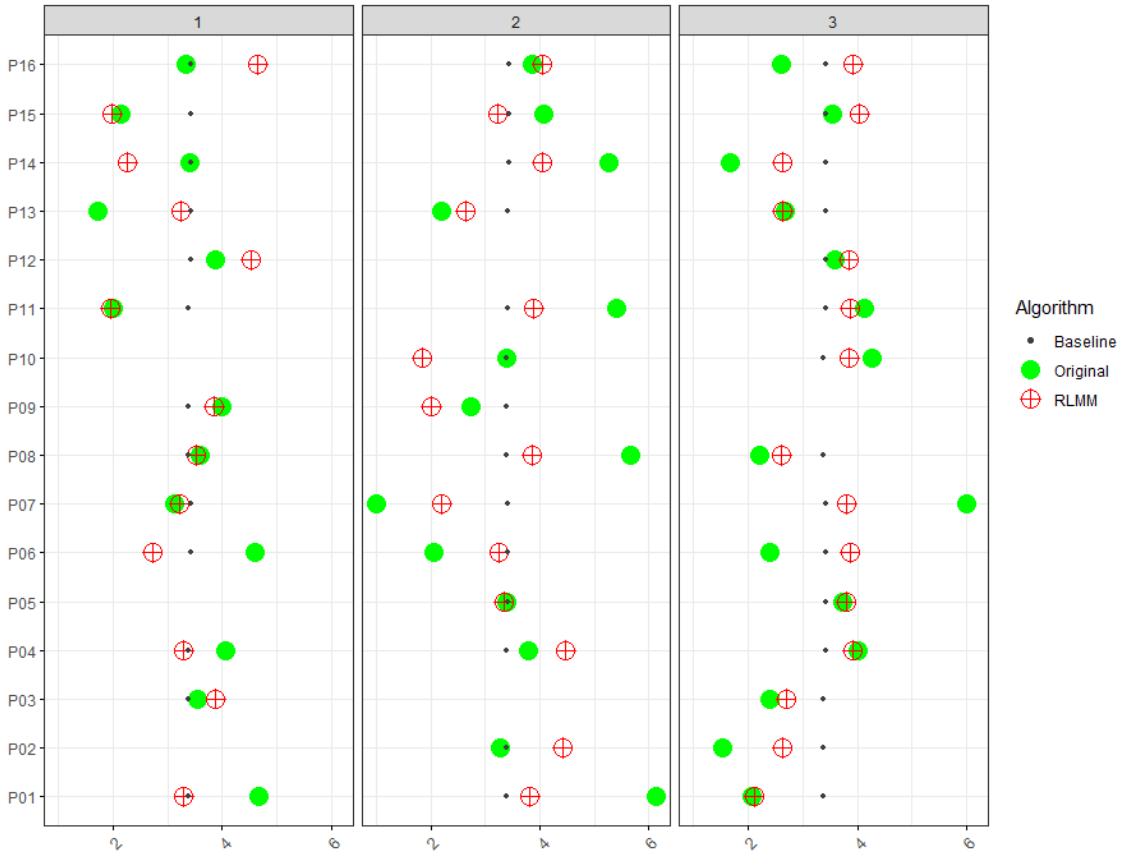



**Figure 25.** Engagement Predictions (based on Random Split) for 42 Observations in Stage 2

### 4.2.3.3.2.4 Comparing Robust LMM with Baseline (One Person Leave Out)

According to the One Person Leave Out evaluation scheme used to predict engagement with both video clickstream behaviors and their reasons, Wilcoxon signed-rank test result shows that Robust LMM (RMSE = 1.09, MAE = 0.86, *SD* = 0.67) did not significantly predict student engagement more accurately than the baseline model (RMSE = 1.12, MAE = 0.91, *SD* = 0.67), *Z* = 507, *p* = .488.

**Table 87.** Comparison of XGBoost, Robust LMM and Baseline for Engagement Prediction (One Person Leave Out in Stage 2)

| *No* | *Model* | *N (observations)* | *%RMSE* | *RMSE* | *MAE* | *SD* | *Z* | *p* |
|---|---|---|---|---|---|---|---|---|
| 1 | Robust LMM | 42 | 31.84 | 1.09 | 0.86 | 0.67 | 507 | .488 |
| 2 | Baseline | 42 | 32.88 | 1.12 | 0.91 | 0.67 | | |

Figure 26 shows predictions for each observation based on the One Person Leave Out scheme with both video clickstream behaviors and their reasons to predict student engagement. The predictions of Robust LMM in this scheme were found to be considerably better than those of the One Person Leave Out scheme with only video clickstream behaviors. However, no significant difference was found to exist between Robust LMM and the baseline model.

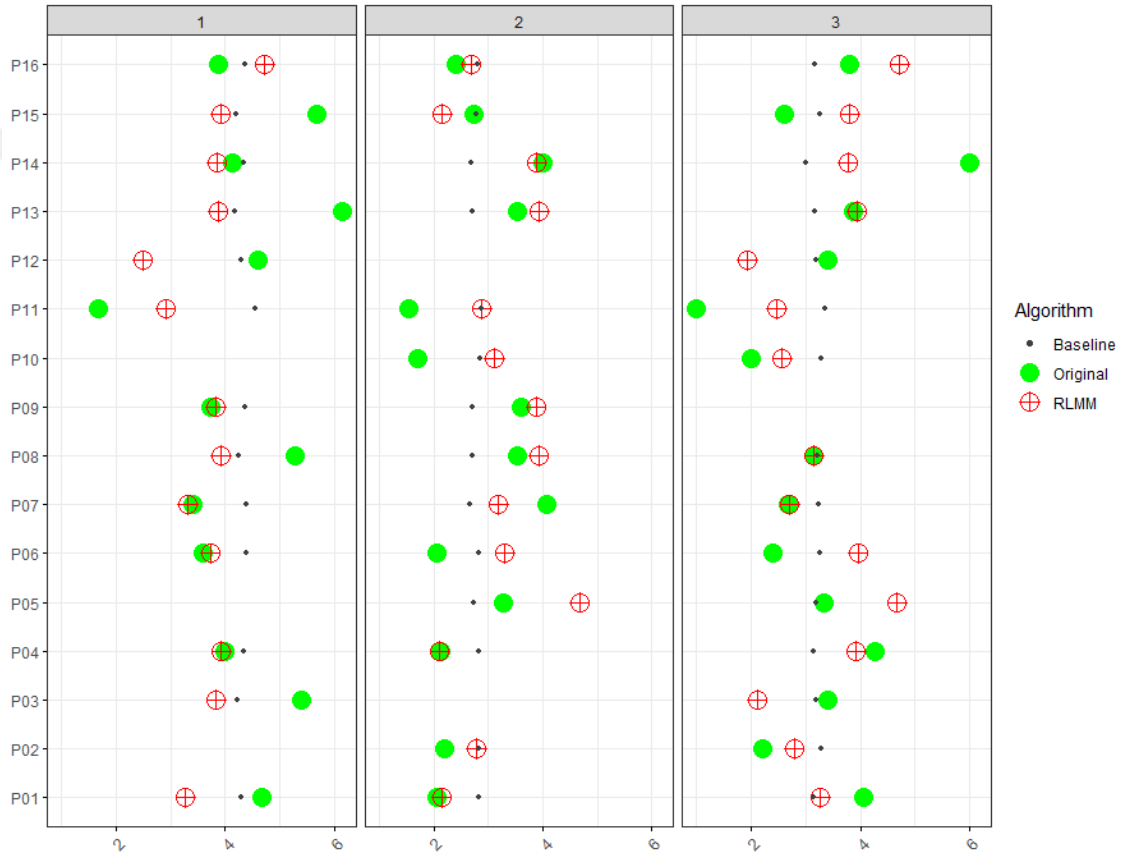



**Figure 26.** Engagement Predictions (based on One Person Leave Out) for 42 Observations in Stage 2

#### 4.2.3.3.2.5 Comparing Robust LMM with Baseline (One Video Leave Out)

According to the One Video Leave Out evaluation scheme used to predict engagement with both video clickstream behaviors and their reasons, the Wilcoxon signed-rank test result showed that Robust LMM (RMSE = 1.19, MAE = 0.95, *SD* = 0.72) did not significantly predict student engagement more accurately than the baseline model (RMSE = 1.28, MAE = 1.04, *SD* = 0.75), $Z = 479$, $p = .731$.

**Table 88.** Comparison of XGBoost, Robust LMM and Baseline for Engagement Prediction (One Video Leave Out in Stage 2)

| *No* | *Model* | *N (observations)* | *%RMSE* | *RMSE* | *MAE* | *SD* | *Z* | *p* |
|---|---|---|---|---|---|---|---|---|
| 1 | Robust LMM | 42 | 34.90 | 1.19 | 0.95 | 0.72 | 479 | .731 |
| 2 | Baseline | 42 | 37.40 | 1.28 | 1.04 | 0.75 | | |

Figure 27 shows the predictions for each observation based on the One Video Leave Out scheme with both video clickstream behaviors and their reasons to predict student engagement. The predictions of Robust LMM in this scheme were found to be considerably better than those for the One Video Leave Out scheme with only video clickstream behaviors. However, no significant difference was found to exist between Robust LMM and the baseline model.

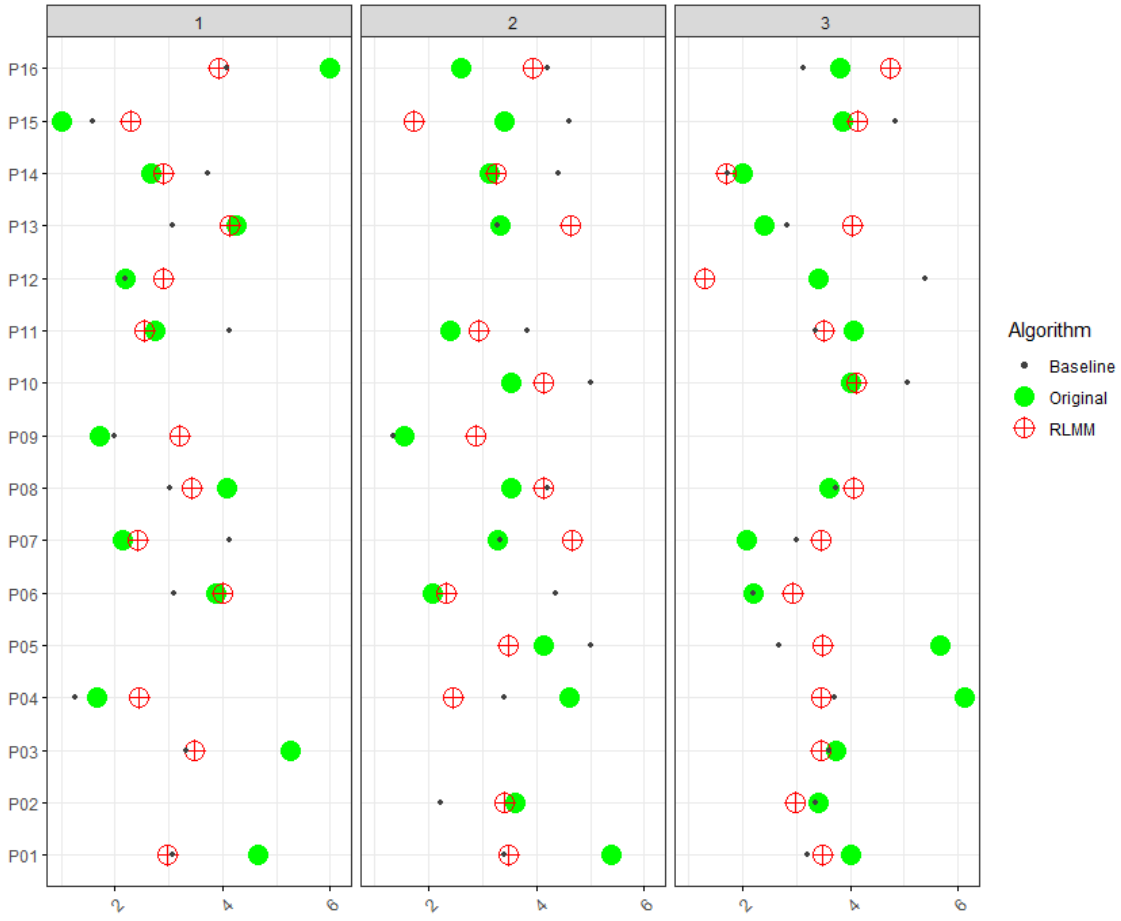



**Figure 27.** Engagement Predictions (based on One Video Leave Out) for 42 Observations in Stage 2

The results based on the three evaluation validation strategies (Random Split, One Person Leave Out, and One Video Leave Out) showed that by using both video clickstream behaviors and their reasons, the selected algorithm did not significantly predict engagement more accurately than the baseline model, which thereby rejects Hypothesis 27.

**Table 89.** Summary of Prediction Findings

| *No* | *Hypothesis* | *Random Split* | | *One Person Leave Out* | | *One Video Leave Out* | | *Accepted* |
|---|---|---|---|---|---|---|---|---|
| | | *%RMSE* | *p* | *%RMSE* | *p* | *%RMSE* | *p* | |
| 23 | *XGBoost* (1) or *Robust LMM* (2) can significantly predict **learning performance** better than the baseline (3) model with "video clickstream behaviors" in terms of three evaluation schemes. | (1) 17.20<br>(2) 17.61<br>(3) 21.12 | .002 | (1) 18.33<br>(2) 17.79<br>(3) 22.26 | .017 | (1) 19.86<br>(2) 21.26<br>(3) 27.79 | .003 | YES |
| 24 | *XGBoost* (1) or *Robust LMM* (2) can significantly predict **satisfaction** better than the baseline (3) model with "video clickstream behaviors" in terms of three evaluation schemes. | (1) 33.03<br>(2) 30.57<br>(3) 31.16 | .146 | (1) 31.85<br>(2) 31.34<br>(3) 29.13 | .909 | (1) 36.04<br>(2) 35.06<br>(3) 27.89 | .254 | NO |
| 25 | The selected algorithm (*Robust LMM)* (1) can significantly predict **satisfaction** better than the baseline (2) model together with "video clickstream behaviors" and "click reasons" in terms of three evaluation schemes. | (1) 18.36<br>(2) 31.16 | .009 | (1) 17.66<br>(2) 29.13 | .042 | (1) 17.07<br>(2) 27.89 | .000 | YES |
| 26 | *XGBoost* (1) or *Robust LMM* (2) can significantly predict **engagement** better than the baseline (3) model with "video clickstream behaviors" in terms of three evaluation schemes. | (1) 37.92<br>(2) 35.73<br>(3) 35.81 | .001 | (1) 41.31<br>(2) 37.26<br>(3) 32.88 | .016 | (1) 41.76<br>(2) 41.69<br>(3) 37.40 | .607 | NO |
| 27 | The selected algorithm (Robust LMM) (1) can significantly predict **engagement** better than the baseline (2) model together with "video clickstream behaviors" and "click reasons" in terms of three evaluation schemes. | (1) 30.26<br>(2) 35.81 | .058 | (1) 31.84<br>(2) 32.88 | .488 | (1) 34.90<br>(2) 37.40 | .731 | NO |

*Quotation marks denote inputs, italics denote algorithms, bold denotes outputs.
*Friedman test results shown for H23, H24, and H25. Multiple comparisons are detailed in Sections 4.2.3.1.1, 4.2.3.2.1, and 4.2.3.3.1.
*Wilcoxon signed-rank test results are shown for H25 and H27.

### 4.3 Threats to Validity

Validity and reliability are important concepts to assuring the quality of research studies. As in every research, the current study also faces certain validity and reliability threats. Some actions and decisions were determined in order to deal with potential threats that may be encountered during or after the study, and the research was designed accordingly.

There are three important types of validity to consider in this type of research: "construct validity," "internal validity," and "external validity."

Construct validity concerns how well a study measures what it intends to measure. In order to measure student satisfaction and engagement in the current study, validated scales were selected. For example, items in the Learner Satisfaction Scale (LSS) exactly and only measured satisfaction, whilst items in the Video Engagement Scale (VES) exactly and only measured engagement. A midterm exam applied during the first experiment (Pilot study) and six quizzes applied during the last experiment (main study) were both prepared and double-checked by the course instructor, who is also a subject-matter expert and has taught the same courses for 20 years. All of the quizzes were created as being of the same level of difficulty in order to answer the research questions during the third experiment. Each quiz included questions related only to the video lecture watched in the previous day. In addition, the interventions (or manipulations) form the interactive elements of the developed framework. For each video, the interventions were each evaluated in terms of their timing and content, and the results presented in Section 3.3.4.1, which demonstrated the use of appropriate interventions based on the developed framework. Moreover, numerous steps were employed in developing the framework. In accordance with the overall purpose, students' opinions and that of an expert were also sought in the development of the framework, and also as a means of ensuring clarity and accuracy in the constructs of the framework.

Internal validity ensures that changes to the dependent variable are due to manipulation of the independent variable. Here there may be two important threats to internal validity: prior knowledge and time. In order to deal with the effect of prior knowledge on the results, only students who were at the same grade level, from the same department, and who were taking the course for the first time were selected as the study's participants. Since the research was based on a one group pretest-posttest design, each student used their same prior knowledge in the pretest-posttest. Time is another potential threat to internal validity because such studies are likely to be affected by the element of time. The measurements taken in this research were collected the day after the students had each watched the allocated videos in order to control the data collection environment, and also to ensure that the measurements were not affected by any other external element or influence. Therefore, the current study also measured the participant students' retention

by way of measuring their learning performance. That is, the students' retention was considered an integral element of learning performance in this research rather than as a manipulator, which demonstrated that time was not a threat posed to the internal validity of the research.

External validity focuses on the generalizability of a study's results. As previously mentioned in Chapter 3, all of the videos selected were approximately 30 minutes in duration. They were otherwise identical, albeit with six different verbalized topics. The results may therefore be considered generalizable for other video lectures of similar length and that also present verbalized topics. However, should the aim be to generalize the results to short videos or quantitative videos, additional research studies would be required.

Reliability concentrates on the replicability of any research. A research protocol outlining the research design was also prepared in order to assure the repeatability of this research. In addition, only scales having high Cronbach's alpha values were used.

# CHAPTER 5

# CONCLUSION

During this research, three different experimental studies were conducted in order to investigate students' video clickstream behaviors, and also to develop a framework based on the behaviors observed. The first experiment was performed as a pilot study with 22 students by using an existing video analytics tool to determine what is needed in a video analytic tool, to gain an idea about potential correlations between students' video watching behaviors and their learning performance, and to identify improvements that could be applied in the subsequent experiment. The second experiment was performed with 29 students using a more comprehensive video analytics tool to develop the framework. Finally, the third experiment was based on the one-group pretest-posttest design, and was conducted as the main study of the research. This main study was conducted with 18 students in order to evaluate the developed framework and to examine the video clickstream behaviors of 16 students having more than one measure.

Based on evaluating the framework, the results showed that the participant students' satisfaction increased significantly when watching interactive videos that were created according to the framework when controlled for motivation. This finding also proves that motivation is an essential factor in student learning (Järvelä, 2001). Even where motivation was not controlled, in using the framework, linear videos could be made interactive without causing any decrease in learning performance.

In addition, the students' *frequency to go back to review important points* was found to decrease significantly in the interactive videos compared to the linear videos, while the students' *frequency to skip unimportant points* increased significantly in the interactive videos. These findings revealed that the framework can be used as a means to decreasing students' cognitive load. That is, students can access the same information through less effort.

The study examined the effect of video clickstream behaviors on learning performance, and showed that the total number of interactions had a significant and positive relationship with the learning performance of the students. That is, the more a student performed clickstream interactions when viewing video lectures, the higher their learning

performance was found to be. This is considered as important evidence of the need for learner-content interaction to be rated successful during online courses (Zimmerman, 2012). The results also showed that a negative correlation exists between *mean of backward speed* and learning performance. It can be said, therefore, that clicking *backward* to some near point in the video being watched demonstrated students wanting to learn, whereas clicking *backward* to some more remote point in a video indicates that the student was not focused on learning.

In order to see group differences, students' video clickstream behaviors were clustered in three different ways; (1) click preferences, (2) number of clicks, and (3) click speed. Clusters were created independently in order to see the specific effect of each clustering input. Clustering based on click preferences revealed four types of groups; (1) *Pauser*, (2) *Backwarder*, (3) *Forwarder*, and (4) *Player*. Clustering based on numbers of clicks revealed two types of groups; (1) *High Clicking* and (2) *Low Clicking*. Clustering based on click speed revealed two types of groups; (1) *Fast Watching* and (2) *Slow Watching*. The clustering results showed that the students achieved greater learning performance when they exhibited:

- *high-clicking* rather than *low-clicking* behavior,
- *backwarder* rather than *other* behavior types, and
- *slow watching* rather than *fast watching* behavior.

The features were then examined so as to predict the students' learning performance, satisfaction, and engagement in a two-stage investigation. The first stage included examining video clickstream behaviors, whereas the second stage considered the reasons behind the students' video watching behaviors. Moreover, XGBoost and Robust Linear Mixed Model (RLMM) algorithms were used to compare the results against baseline models. The results were validated using three different evaluation schemes: Random Split, One Person Leave Out, and One Video Leave Out.

From all three evaluation schemes, both XGBoost and RLMM were shown to more accurately predict the students' learning performance from using their video clickstream behaviors when compared to the baseline model. On the other hand, RLMM predicted satisfaction more accurately than the baseline model together with video clickstream behaviors and their reasons in terms of three evaluation schemes.

The most important indicators predicting learning performance were shown to be *number of pauses*, *mean of backward speed*, and *number/percentage of backwards*. These indicators are compatible with the findings reported in the literature, even though none of them have been clearly expressed for the prediction of learning performance in any single study. Lan et al. (2017) showed that the number of pauses and rewinds were among the features that contribute to students' learning the most, which shows consistency with the

findings of the current study. Furthermore, Hasan et al. (2020) indicated that pausing videos many times or progressing slowly through videos increases the probability of successfully passing a course. Lu et al. (2018) found that the number of backwards clicks per week was also a relevant factor in predicting learning performance. This also justifies the findings in the current study.

For predicting satisfaction, the most important indicators were *frequency of wanting to finish quickly*, *number of backwards*, *number of pauses*, *frequency of facing language issues*, *mean of forward speed*, *frequency of notetaking*, *frequency of going back to review important points*, *mean of backward speed*, and *percentage of pauses*. However, it was not found possible to predict engagement to a statistically significant level using video clickstream behaviors and their reasons.

In addition, it may be speculated that students' learning performance and satisfaction can be predicted with an error in the range of approximately 15 to 20 points (%RMSE) over a 100-point range. As stated by Li et al. (2013), a 15-20% RMSE value is considered good in predicting students' learning performance or satisfaction though examining video clickstream interactions. In other words, learning performance can be predicted to a high percentage of accuracy by using only video clickstream behaviors, and that satisfaction can be predicted by using both video clickstream behaviors and their reasons. However, information additional to just video clickstream behaviors and their reasons are needed in order to predict engagement. Hu and Li (2017) also emphasizes that it is not sufficient to assess learners' engagement only from learners' behaviors, thereby some qualitative data is needed to make predictions about engagement.

In conclusion, the current research revealed important results in terms of both the framework and comprehensive learning analytics. The results related to the developed framework may be summarized as follows: The framework can decrease students' cognitive load since it changes their behavior in terms of *frequency of skipping unimportant points* and *frequency of going back to review important points* in a video being viewed. The framework can therefore be used to help satisfy students viewing video lectures if their motivation is controlled. It can also be used for transforming linear videos into interactive videos. Moreover, the framework can help to establish a baseline for the building of smart and adaptable e-learning systems.

In terms of comprehensive learning analytics, the study's results may be summarized as follows; Students performing *high clicking* behavior, *slow backward* behavior, or preferring *backwarder* behavior realize higher learning performance results than those students performing *low clicking* behavior, *fast forward* behavior, or preferring *other* behavior types. In addition, the study showed that video clickstream behaviors are sufficient to predict the learning performance of students; however, the reasons behind the behaviors are also necessary to predict satisfaction in students in addition to the behaviors,

whilst yet more information is deemed necessary for engagement to be accurately predicted.

### 5.1 Implications for Learning Design and Learning Experience Design

Learning analytics utilize learners' interactions in order to reveal and better understand their learning experiences (Hernández-Leo et al., 2019). Similarly, the term "learning experiences" encompasses learning interactions within a learning environment (Şahin, 2018). In addition, learning design focuses upon being learner-centered in order to support learning experiences. Many studies have attempted to link learning design and learning analytics (Mangaroska & Giannakos, 2018; Toetenel & Rienties, 2016); however, the results published for video-based learning analytics have yet to be associated with learning design. Therefore, this section attempts to separately associate the findings of the current study with both Learning Experience Design (LXD) and Learning Design (LD). Both of these design approaches concentrate on the learner, but each uses different concepts even though both have been used interchangeably in numerous studies.

First, design suggestions based on the results of the current study may be used to help increase the value of students' learning experiences from watching educational videos and lectures. There are four important perspectives that a learning experience designer should focus upon: Goal Oriented, Human Oriented, Theory of Learning, Learning Put into Practice (Floor, 2019; Sousa & Martins, 2021).

- Goal oriented: to address a specific purpose.
- Human oriented: to focus on the learners' needs and requests.
- Theory of learning: grounded upon proven scientific theory.
- Learning put into practice: to enable learners to practice what they learn.

Since the primary goal is to help students learn, video lectures should be designed in such a way that students perform behaviors that result in or lead to higher learning performance. Therefore, following some of the current study's findings can be considered as expediting the design and production of goal-oriented videos:

- Learning performance is higher for *Backwarders* rather than *others.*
- Learning performance is higher with *High Clicking* rather than *Low Clicking.*
- Learning performance is higher with *Slow Watching* rather than *Fast Watching.*

In addition, the embedding of questions into video lectures can help improve learning, as well as the amount of time spent viewing video lectures (Vural, 2013). Embedded questions can also help contribute to a mental review or replay of video lectures in order to achieve a learning goal (Van der Meij & Böckmann, 2020). Therefore, in order to increase learners' *backwarder*, *high clicking*, and *slow watching* behaviors, **interactive questions (for *high clicking*) that provide students with stimuli to search for and**

**process information (for *backwarder behavior*) should be added/embedded periodically to video lectures (for *slow watching*)**. This approach can also help students in their aspirations to learn by enabling them to put into practice what they learn through interactive and inquisitive questioning.

In order to provide human-oriented educational videos, **motivational interventions should be increased or changed/enhanced** since students mostly need to be actively motivated, and research results, such as seen in the current study, have shown the importance and positive effect of students' motivation on their learning. Şahin (2018) also emphasized that motivational interventions can be performed as a result of learning experiences.

Video lectures should be designed based on a proven theory of learning. Since the current study included results regarding students' cognitive load, and many educators have emphasized the importance of cognitive load theory, the established theories of learning can be utilized to create more effective video lectures. Cognitive load theory focuses on four different practices: signaling, segmenting, weeding, and matching modality (Brame, 2016). First, signaling refers to highlighting essential parts or cues in multimedia learning (Mayer, 2009). Placing text or images referring to important points when creating interactive videos is a part of signaling and the act of going back to review important points significantly decreases with interactive videos. Therefore, **text or images should be added to important points in video lectures**. Second, segmenting means the breaking down of multimedia into smaller units (Ibrahim et al., 2012). In terms of segmenting, **video durations should be kept short** or **videos should be divided into smaller parts**. When the duration of a video is kept appropriately short, students do not the need to watch the videos too quickly, which therefore can lead to more effective "slow watching" rather than "fast watching." Third, weeding implies the removal of irrelevant parts (Mayer & Moreno, 2003), and is usually around **the points from where students tend to skip forward** and therefore **should be summarized in order to decrease *forwarder* behavior** (Özhan et al., 2019) or **to decrease *forward speed*** which is an important indicator for both student satisfaction and engagement. All of these are also consistent with Mayer (2014)'s cognitive theory on multimedia learning as they are about both video lectures and cognitive load.

In addition to learning experience design, learning design activities should also be taken into account. Learning design consists of seven different activity types (Rienties & Toetenel, 2015): assessment, assimilative, communication, experiential, finding and handling information, interactive, and productive. Using in-video interventions when creating interactive videos helps contribute to the learning design, especially in terms of interactive, experiential, and assessment activities in video lectures. According to Holmes et al. (2019), students achieving the highest learning performance often belong to those having received the highest number of assessment activities and the lowest number of

assimilative activities (e.g., reading, watching, listening, etc.). **This combination can be facilitated by increasing the number of assessment questions embedded into a video lecture, and/or decreasing its duration.** In-video assessments also increase interactivity and formative assessment (Cummins et al., 2016). Therefore, using the framework developed in the current study together with the optimum number of questions in short instructional videos would likely have affected the study's results. The study also showed that students have the highest satisfaction when the number of interactive activities is the highest, even though no statistically significant difference was found between the groups. Therefore, considering these points, the developed framework could be more effective in video lectures if the optimum conditions such as video duration and number of questions are met.

In summary, an educator should consider the following recommendations when designing video lectures in order to maximize their students' learning. As such, the following key highlights should be observed:

- Keep video lectures as short as possible.
- Increase in-video assessment activities.
- Add text or images to the important points within video lectures.
- Improve motivational interventions.
- Employ interactive questioning in video lectures that provide learners with the opportunity to periodically search for and process information.
- If previous recommendations made about a video are unavailable, at least summarize the points where students tend to skip forward.

### 5.2 Limitations and Future Work

Even though the current study provides valuable insights into a framework based on the video clickstream behaviors of students, and also presents clusters, predictors, and correlations related with video clickstream behaviors, it does present certain limitations as a piece of research.

First, the study included only a small number of participants. One of the most important reasons for this was that three different experimental studies with repeated measures were performed, which therefore took up considerable time for each participant. With a limited number of available participants, the sample could not have been assigned to two different groups, such as with a control group and an experimental group. Therefore, a quasi-experimental study was conducted as the main study (Experiment 3) with a pretest-posttest design. Second, different tools were used for the linear and interactive videos due

to technical limitations. Therefore, the students' video viewing behaviors based on clickstream interactions were not tracked in the case of the three interactive videos.

As a future piece of work, experimental studies on different courses are planned in order to generalize the framework for different course types, and thereby to observe and assess their differences. The developed framework may also help to address and diminish the achievement gap between groups of students; therefore, further research is also planned that will aim to measure the effect of the framework on any achievement gap between students.

## REFERENCES


Adnan, H., & Redzuan, F. (2017). Evaluating students' emotional response in video-based learning using Kansei Engineering. In *Proceedings 2016 4th International Conference on User Science and Engineering, i-USEr* (pp. 237-242). IEEE. https://doi.org/10.1109/IUSER.2016.7857967

Aguiar, E., Nagrecha, S., & Chawla, N. V. (2015). Predicting online video engagement using clickstreams. In *Proceedings of the 2015 IEEE International Conference on Data Science and Advanced Analytics, DSAA 2015*. IEEE. https://doi.org/10.1109/DSAA.2015.7344873

Akçapınar, G., & Bayazıt, A. (2018). Investigating Video Viewing Behaviors of Students with Different Learning Approaches. *Turkish Online Journal of Distance Education, 19*(4), 116-125. https://doi.org/10.17718/tojde.471907

Akinwande, M. O., Dikko, H. G., & Samson, A. (2015). Variance Inflation Factor: As a Condition for the Inclusion of Suppressor Variable(s) in Regression Analysis. *Open Journal of Statistics, 05*(07), 754-767. https://doi.org/10.4236/ojs.2015.57075

Al-Azawei, A., & Lundqvist, K. (2015). Learner differences in perceived satisfaction of an online learning: An extension to the technology acceptance model in an Arabic sample. *Electronic Journal of E-Learning, 13*(5), 408-426. https://academic-publishing.org/index.php/ejel/article/view/1942

Al-Azawei, A., Parslow, P., & Lundqvist, K. (2017). Investigating the effect of learning styles in a blended e-learning system: An extension of the technology acceptance model (TAM). *Australasian Journal of Educational Technology, 33*(2), 1-23. https://doi.org/10.14742/ajet.2741

Allen, M., Bourhis, J., Burrell, N., & Mabry, E. (2002). Comparing Student Satisfaction With Distance Education to Traditional Classrooms in Higher Education: A Meta-Analysis. *International Journal of Phytoremediation, 21*(1), 83-97. https://doi.org/10.1207/S15389286AJDE1602_3

Almeda, M. V., Zuech, J., Baker, R. S., Utz, C., Higgins, G., & Reynolds, R. (2018). Comparing the Factors that Predict Completion and Grades among For-Credit and Open/MOOC students in Online Learning. *Online Learning, 22*(1). http://dx.doi.org/10.24059/olj.v22i1.1060

Al-rahmi, W. M., Othman, M. S., & Yusuf, L. M. (2015). Using Social Media for Research : The Role of Interactivity , Collaborative Learning, and Engagement on the Performance of Students in Malaysian Post-Secondary Institutes. *Mediterranean Journal of Social Sciences*, *6*(5), 536-546. https://doi.org/10.5901/mjss.2015.v6n5s2p536

Bakharia, A., Corrin, L., Barba, P. De, Kennedy, G., Gasevic, D., Mulder, R., Williams, D., Dawson, S., & Lockyer, L. (2016). A Conceptual Framework linking Learning Design with Learning Analytics. In *Proceedings of LAK16 6th International Conference on Analytics and Knowledge* (pp. 329-338). ACM. https://doi.org/10.1145/2883851.2883944

Bakla, A. (2017). Interactive Videos in Foreign Language Instruction: A New Gadget in Your Toolbox. *Mersin University Journal of the Faculty of Education*, *13*(1), 124-137. http://dx.doi.org/10.17860/mersinefd.305769

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*(1), 289-300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Blasco-arcas, L., Buil, I., Hernández-ortega, B., & Sese, F. J. (2013). Using clickers in class. The role of interactivity, active collaborative learning and engagement in learning performance. *Computers & Education*, *62*, 102-110. https://doi.org/10.1016/j.compedu.2012.10.019

Brame, C. J. (2016). Effective educational videos: Principles and guidelines for maximizing student learning from video content. *CBE Life Sciences Education*, *15*(4), es6.1-es6.6. https://doi.org/10.1187/cbe.16-03-0125

Brinton, C. G., Buccapatnam, S., Chiang, M., & Poor, H. V. (2015). *Mining MOOC Clickstreams : On the Relationship Between Learner Video-Watching Behavior and Performance*. https://arxiv.org/abs/1503.06489

Brinton, C. G., Buccapatnam, S., Chiang, M., & Poor, H. V. (2016). Mining MOOC Clickstreams: Video-Watching Behavior vs. In-Video Quiz Performance. *IEEE Transactions on Signal Processing*, *64*(14), 3677-3692. https://doi.org/10.1109/TSP.2016.2546228

Brooks, C., Thompson, C., & Greer, J. (2013). Visualizing lecture capture usage: A learning analytics case study. In M. N. Giannakos, K. Chorianopoulos, M. Ronchetti, P. Szegedi, & S. Teasley (Eds.), *Proceedings of the LAK 2013 Workshop on Analytics on Video-based Learning* (Vol. 983, pp. 9-14). ACM. http://ceur-ws.org/Vol-983/paper3.pdf

Buelow, J. R., Barry, T., & Rich, L. E. (2018). Supporting learning engagement with online students. *Online Learning Journal*, *22*(4), 313-340. https://doi.org/10.24059/olj.v22i4.1384

Chatti, M. A., Marinov, M., Sabov, O., Laksono, R., Sofyan, Z., Fahmy Yousef, A. M., & Schroeder, U. (2016). Video annotation and analytics in CourseMapper. *Smart Learning Environments*, *3*(1), Article 10. https://doi.org/10.1186/s40561-016-0035-1

Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785-794). ACM. https://doi.org/10.1145/2939672.2939785

Chen, T., He, T., & Benesty, M. (2015). *xgboost: eXtreme Gradient Boosting. R Package Version 0.4-0, 0–4.* https://cran.microsoft.com/snapshot/2015-10-20/web/packages/xgboost/xgboost.pdf

Cherrett, T., Wills, G., Price, J., Maynard, S., & Dror, I. E. (2009). Making training more cognitively effective: Making videos interactive. *British Journal of Educational Technology*, *40*(6), 1124–1134. https://doi.org/10.1111/j.1467-8535.2009.00985.x

Cho, M., & Cho, Y. (2014). Instructor scaffolding for interaction and students' academic engagement in online learning : Mediating role of perceived online class goal structures. *The Internet and Higher Education*, *21*, 25-30. https://doi.org/10.1016/j.iheduc.2013.10.008

Cho, T. J. (2011). The Impact of Types of Interaction on Student Satisfaction in Online Courses. *International Journal on E-Learning*, *10*(2), 109-125. https://www.learntechlib.org/p/32381/

Choe, R. C., Scuric, Z., Eshkol, E., Cruser, S., Arndt, A., Cox, R., Toma, S. P., Shapiro, C., Levis-Fitzgerald, M., Barnes, G., & Crosbie, R. H. (2019). Student satisfaction and learning outcomes in asynchronous online lecture videos. *CBE Life Sciences Education*, *18*(4), 1-14. https://doi.org/10.1187/cbe.18-08-0171

Chorianopoulos, K., Giannakos, M. N., Chrisochoides, N., & Reed, S. (2014). Open service for video learning analytics. In *Proceedings - IEEE 14th International Conference on Advanced Learning Technologies* (pp. 28-30). IEEE. https://doi.org/10.1109/ICALT.2014.19

Coldwell, J., Craig, A., Paterson, T., & Mustard, J. (2008). Online students: Relationships between participation, demographics and academic performance. *Electronic Journal of E-Learning*, *6*(1), 19–28. https://academic-publishing.org/index.php/ejel/article/view/1528

Conijn, R., Van den Beemt, A., & Cuijpers, P. (2018). Predicting student performance in a blended MOOC. *Journal of Computer Assisted Learning*, *34*(5), 615-628. https://doi.org/10.1111/jcal.12270

Conole, G., Klobučar, T., Rensing, C., Konert, J., & Lavoué, É. (2015). MOOC Video Interaction Patterns: What Do They Tell Us? *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 9307*, 197-210. https://doi.org/10.1007/978-3-319-24258-3

Cummins, S., Beresford, A. R., & Rice, A. (2016). Investigating Engagement with In-Video Quiz Questions in a Programming Course. *IEEE Transactions on Learning Technologies*, *9*(1), 57-66. https://doi.org/10.1109/TLT.2015.2444374

Czerkawski, B. C., & Lyman, E. W. (2016). An Instructional Design Framework for Fostering Student Engagement in Online Learning Environments. *TechTrends*, *60*(6), 532–539. https://doi.org/10.1007/s11528-016-0110-z

Daniels, L. M., Adams, C., & McCaffrey, A. (2016). Emotional and Social Engagement in a Massive Open Online Course. *Emotions, Technology, and Learning*, *2016*, 25-41. https://doi.org/10.1016/b978-0-12-800649-8.00004-3

DeBourgh, G. A. (2003). Predictors of student satisfaction in distance-delivered graduate nursing courses: What matters most? *Journal of Professional Nursing*, *19*(3), 149-163. https://doi.org/10.1016/S8755-7223(03)00072-3

Demidenko, E. (2013). *Mixed Models : Theory and Applications with R* (2nd ed.). Wiley.

Dennen, V. P., Darabi, A. A., & Smith, L. J. (2007). Instructor-learner interaction in online courses: The relative perceived importance of particular instructor actions on performance and satisfaction. *Distance Education*, *28*(1), 65-79. https://doi.org/10.1080/01587910701305319

Dimitrova, V., Mitrovic, A., Piotrkowicz, A., Lau, L., & Weerasinghe, A. (2017). Using Learning Analytics to Devise Interactive Personalised Nudges for Active Video Watching. In *Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization - UMAP '17* (pp. 22-31). ACM. https://doi.org/10.1145/3079628.3079683

Dogan, U. (2015). Student engagement, academic self-efficacy, and academic motivation as predictors of academic performance. *Anthropologist*, *20*(3), 553-561. https://doi.org/10.1080/09720073.2015.11891759

Donkor, F. (2011). Assessment of learner acceptance and satisfaction with video-based instructional materials for teaching practical skills at a distance. *International Review of Research in Open and Distance Learning*, *12*(5), 71-88. https://doi.org/10.19173/irrodl.v12i5.953

Dzakira, H. (2003). Teacher- Learner Interactions in Distance Education: A Case of Two Malaysian Universities. *The Turkish Online Journal of Distance Education*, *4*(3). https://dergipark.org.tr/en/pub/tojde/issue/16937/176820

Faraway, J. J. (2002). *Practical Regression and Anova using R*. https://cran.microsoft.com/snapshot/2015-03-01/doc/contrib/Faraway-PRA.pdf

Fernández-Gallego, B., Lama, M., Vidal, J. C., & Mucientes, M. (2013). Learning analytics framework for educational virtual worlds. *Procedia Computer Science*, *25*, 443-447. https://doi.org/10.1016/j.procs.2013.11.056

Field, A. P. (2018). *Discovering Statistics Using IBM SPSS Statistics* (5th ed.). Sage.

Field, A., & Hole, G. (2003). *How to Design and Report Experiments*. Sage.

Floor, N. (2019). *The origin of learning experience design*. https://lxd.org/fundamentals-of-learning-experience-design/the-origin-of-learning-experience-design/

Fox, J. (2002). Linear mixed models for R (notes). *Statistics in Medicine*, *29*(25), 2643-2655. http://www.ncbi.nlm.nih.gov/pubmed/20740568

Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2012). *How to Design and Evaluate Research in Education* (8th ed.). McGraw Hill.

Gameel, B. G. (2017). Learner Satisfaction with Massive Open Online Courses. *American Journal of Distance Education*, *31*(2), 98-111. https://doi.org/10.1080/08923647.2017.1300462

Gardner, J., & Brooks, C. (2018). Student success prediction in MOOCs. *User Modeling and User-Adapted Interaction*, *28*(2), 127-203. https://doi.org/10.1007/s11257-018-9203-z

Geller, E. S. (2005). Behavior-Based Safety and Occupational Risk Management. *Behavior Modification*, *29*(3), 539-561. https://doi.org/10.1177/0145445504273287

Ghaemi, R., & Lam, G. (2021). Characterizing Patterns in Student Engagement in A Team-Based and Project-Centric Course. In *Proceedings 2021 Canadian Engineering Education Association (CEEA-ACEG21) Conference*. CEEA. https://doi.org/10.24908/pceea.vi0.14921

Giannakos, M. N., Chorianopoulos, K., & Chrisochoides, N. (2015). Collecting and making sense of video learning analytics. In *Proceedings - Frontiers in Education Conference, FIE, 2015*. IEEE. https://doi.org/10.1109/FIE.2014.7044485

Giannakos, M. N., Chorianopoulos, K., Ronchetti, M., Szegedi, P., & Teasley, S. D. (2013). Expanding horizons and envisioning the future of analytics on video-based learning. In M. N. Giannakos, K. Chorianopoulos, M. Ronchetti, P. Szegedi, & S. Teasley (Eds.), *Proceedings of the LAK 2013 Workshop on Analytics on Video-based Learning* (Vol. *983*, pp. 1-6). CEUR. http://ceur-ws.org/Vol-983/paper1.pdf

Giannakos, M. N., Chorianopoulos, K., Ronchetti, M., Szegedi, P., & Teasley, S. D. (2014). Video-Based Learning and Open Online Courses. *International Journal of Emerging Technologies in Learning (IJET)*, *9*(1), 4-7. https://doi.org/10.3991/ijet.v9i1.3354

Giannakos, M. N., Sampson, D. G., Kidziński, Ł., & Pardo, A. (2016). Enhancing video-based learning experience through smart environments and analytics. In M. N. Giannakos, D. G. Sampson, Ł. Kidziński, & A. Pardo (Eds.), *Proceedings of the LAK 2016 Workshop on Smart Environments and Analytics in Video-Based Learning* (Vol. 1579, pp. 26-33). CEUR. http://ceur-ws.org/Vol-1579/paper5.pdf

Gong, L., & Liu, Y. (2019). Design and Application of Intervention Model based on Learning Analytics under Blended Learning Environment. In *Proceedings of the 2019 7th International Conference on Information and Education Technology (ICET 2019)* (pp. 225-229). ACM. https://doi.org/10.1145/3323771.3323825

Greller, W., & Drachsler, H. (2012). Translating learning into numbers: A generic framework for learning analytics. *Educational Technology and Society*, *15*(3), 42-57. https://drive.google.com/file/d/1R84FXoT3W3X6C2JV1BBXha3tCoOQiQ7l/view

Guo, B., Zhang, R., Xu, G., Shi, C., & Yang, L. (2015). Predicting Students Performance in Educational Data Mining. In F. L. Wang, T. L. Wong, O. Au, Q. Liu, & D. Wu (Eds.), *Proceedings 2015 International Symposium on Educational Technology (ISET)* (pp. 127-130). IEEE. https://doi.org/10.1109/ISET.2015.33

Guo, P. J., Kim, J., & Rubin, R. (2014). How video production affects student engagement: An empirical study of MOOC videos. In *Proceedings of the 1st ACM Conference on Learning at Scale - L@S 2014* (pp. 41-50). ACM. https://doi.org/10.1145/2556325.2566239

Gupta, C. P., Singh, B., & Marwaha, T. (2013). Relationship between Social Media and Academic Performance in Distance Education. *Universal Journal of Educational Research*, *1*(3), 185-190. https://doi.org/10.13189/ujer.2013.010307

Guay, F., Vallerand, R. J., & Blanchard, C. (2000). On the Assessment of Situational Intrinsic and Extrinsic Motivation: The Situational Motivation Scale. *Motivation and Emotion*, *24*(3), 175–213. https://doi.org/https://doi.org/10.1023/A:1005614228250

Harandi, S. R. (2015). Effects of e-learning on Students' Motivation. *Procedia - Social and Behavioral Sciences*, *181*, 423–430. https://doi.org/10.1016/j.sbspro.2015.04.905

Hasan, R., Palaniappan, S., Mahmood, S., & Abbas, A. (2020). Predicting Student Performance in Higher Educational Institutions Using Video Learning Analytics and Data Mining Techniques. *Applied Science*, *10*(11), Article 3894. https://doi.org/10.3390/app10113894

Hasnine, M. N., Flanagan, B., & Majumdar, R. (2018). *Towards Final Scores Prediction over Clickstream Using Machine Learning Methods*. *November*.

Hernández-Leo, D., Martinez-Maldonado, R., Pardo, A., Muñoz-Cristóbal, J. A., & Rodríguez-Triana, M. J. (2019). Analytics for learning design: A layered framework and tools. *British Journal of Educational Technology*, *50*(1), 139-152. https://doi.org/10.1111/bjet.12645

Hew, K. F., Hu, X., Qiao, C., & Tang, Y. (2020). What predicts student satisfaction with MOOCs: A gradient boosting trees supervised machine learning and sentiment analysis approach. *Computers and Education*, *145*, Article 103724. https://doi.org/10.1016/j.compedu.2019.103724

Hildebrand, J. D., & Ahn, B. (2018). Student video viewing habits in an online mechanics of materials engineering course. *International Journal of Engineering Pedagogy*, *8*(3), 40-59. https://doi.org/10.3991/ijep.v8i3.7948

Hilgenberg, C., & Tolone, W. (2000). Student perceptions of satisfaction and opportunities for critical thinking in distance education by interactive video. *International Journal of Phytoremediation*, *21*(1), 59-73. https://doi.org/10.1080/08923640009527065

Ho, A. D., Reich, J., Nesterko, S. O., Seaton, D. T., Mullaney, T., Waldo, J., & Chuang, I. (2014). HarvardX and MITx: The First Year of Open Online Courses, Fall 2012-Summer 2013. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2381263

Hollands, F., & Kazi, A. (2018). *Benefits and Costs of MOOC-Based Alternative Credentials (2018).* Center for Benefit-Cost Studies of Education, Teachers College, Columbia University. . https://www.academia.edu/37862783/Benefits_and_Costs_of_MOOC_Based_Alternative_Credentials_2018_

Holmes, W., Nguyen, Q., Zhang, J., Mavrikis, M., & Rienties, B. (2019). Learning analytics for learning design in online distance learning. *Distance Education*, *40*(3), 309-329. https://doi.org/10.1080/01587919.2019.1637716

Hu, H., Zhang, G., Gao, W., & Wang, M. (2020). Big data analytics for MOOC video watching behavior based on Spark. *Neural Computing and Applications*, *32*(11), 6481-6489. https://doi.org/10.1007/s00521-018-03983-z

Hu, M., & Li, H. (2017). Student engagement in online learning: A review. In F. L. Wang, O. Au, K. K. Ng, J. Shang, & R. Kwan (Eds.), *Proceedings 2017 International Symposium on Educational Technology, ISET 2017* (pp. 39-43). IEEE. https://doi.org/10.1109/ISET.2017.17

Ibrahim, M., Antonenko, P. D., Greenwood, C. M., & Wheeler, D. (2012). Effects of segmenting, signalling, and weeding on learning from educational video. *Learning, Media and Technology*, *37*(3), 220-235. https://doi.org/10.1080/17439884.2011.585993

Jamieson, P. D., Porter, J. R., & Wilson, D. R. (1991). A test of the computer simulation model ARCWHEAT1 on wheat crops grown in New Zealand. *Field Crops Research*, *27*(4), 337-350. https://doi.org/10.1016/0378-4290(91)90040-3

Järvelä, S. (2001). Shifting Research on Motivation and Cognition to an Integrated Approach on Learning and Motivation in Context. In *Motivation in Learning Contexts: Theoretical Advances and Methodological Implications* (pp. 3–14). Pergamon/Elsevier.

Järvelä, S., Järvenoja, H., Malmberg, J., Isohätälä, J., & Sobocinski, M. (2016). How do types of interaction and phases of self-regulated learning set a stage for collaborative engagement? *Learning and Instruction*, *43*, 39-51. https://doi.org/10.1016/j.learninstruc.2016.01.005

Jeong, H., Jang, T., Seong, C., & Park, S. (2014). Assessing nitrogen fertilizer rates and split applications using the DSSAT model for rice irrigated with urban wastewater. *Agricultural Water Management*, *141*, 1-9. https://doi.org/10.1016/j.agwat.2014.04.009

Jung, Y., & Lee, J. (2018). Learning Engagement and Persistence in Massive Open Online Courses (MOOCS). *Computers and Education*, *122*, 9-22. https://doi.org/10.1016/j.compedu.2018.02.013

Kahn, P., Everington, L., Kelm, K., Reid, I., & Watkins, F. (2017). Understanding student engagement in online learning environments: the role of reflexivity. *Educational Technology Research and Development*, *65*(1), 203-218. https://doi.org/10.1007/s11423-016-9484-z

Keeler, L. C. (2006). *Student Satisfaction And Types Of Interaction In Distance Education Courses* [Doctoral dissertation, Colorado State University]. https://www.proquest.com/openview/b574767d2c870ad758dfb550d2889f56/1?pq-origsite=gscholar&cbl=18750&diss=y

Khalil, H., & Ebner, M. (2015). "How Satisfied Are You With Your MOOC?"—A Research Study About Interaction in Huge Online Courses. *Journalism and Mass Communication, 5*(12). https://doi.org/10.17265/2160-6579/2015.12.003

Kim, J., Guo, P. J., Cai, C. J., Li, S. W., Gajos, K. Z., & Miller, R. C. (2014). Data-driven interaction techniques for improving navigation of educational videos. In *UIST 2014 - Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology* (pp. 563-572). ACM. https://doi.org/10.1145/2642918.2647389

Kim, J., Li, S.-W., Cai, C. J., Gajos, K. Z., & Miller, R. C. (2014, April 26 - May 1). *Leveraging Video Interaction Data and Content Analysis to Improve Video Learning*. CHI '14 Workshop on Learning Innovation at Scale, Toronto, Canada.

Kleftodimos, A., & Evangelidis, G. (2014a). Exploring student viewing behaviors in online educational videos. In *Proceedings - IEEE 14th International Conference on Advanced Learning Technologies, ICALT 2014* (pp. 367-369). IEEE. https://doi.org/10.1109/ICALT.2014.109

Kleftodimos, A., & Evangelidis, G. (2014b). Using metrics and cluster analysis for analyzing learner video viewing behaviours in educational videos. In A. Bouras, Z. Tari, A. Erradi, & S. Abdelwahed (Eds.), *Proceedings of IEEE/ACS International Conference on Computer Systems and Applications, AICCSA, 2014* (pp. 280-287). IEEE. https://doi.org/10.1109/AICCSA.2014.7073210

Kleftodimos, A., & Evangelidis, G. (2016a). An interactive video-based learning environment that supports learning analytics for teaching "Image Editing." In M. N. Giannakos, D. G. Sampson, Ł. Kidziński, & A. Pardo (Eds.), *Proceedings of the LAK 2016 Workshop on Smart Environments and Analytics in Video-Based Learning* (Vol. 1579, pp. 26-33). CEUR. http://ceur-ws.org/Vol-1579/paper5.pdf

Kleftodimos, A., & Evangelidis, G. (2016b). An Interactive Video-Based Learning Environment Supporting Learning Analytics: Insights Obtained from Analyzing Learner Activity Data. In Y. Li, M. Chang, M. Kravcik, E. Popescu, R. Huang, Kinshuk, & N.-S. Chen (Eds.), *State-of-the-Art and Future Directions of Smart Learning* (pp. 471-481). Springer. https://doi.org/10.1007/978-981-287-868-7

Klein, C., Lester, J., Nguyen, T., Justen, A., Rangwala, H., & Johri, A. (2019). Student Sensemaking of Learning Analytics Dashboard Interventions in Higher Education. *Journal of Educational Technology Systems*, *48*(1), 130-154. https://doi.org/10.1177/0047239519859854

Koller, M. (2016). Robustlmm: An R package for Robust estimation of linear Mixed-Effects models. *Journal of Statistical Software*, *75*(6). https://doi.org/10.18637/jss.v075.i06

Körösi, G., Esztelecki, P., Farkas, R., & Toth, K. (2018). Clickstream-based outcome prediction in short video MOOCs. In M. S. Obiadat, P. Lorenz, K.-F. Hsiao, P. Nicopolitidis, & D. Cascado-Caballero (Eds.), *Proceedings - 2018 International Conference on Computer, Information and Telecommunication Systems (CITS 2018).* IEEE. https://doi.org/10.1109/CITS.2018.8440182

Kuo, Y.-C., Walker, A. E., Schroder, K. E. E., & Belland, B. R. (2014). Interaction, Internet self-efficacy, and self-regulated learning as predictors of student satisfaction in online education courses. *Internet and Higher Education, 20,* 35-50. https://doi.org/10.1016/j.iheduc.2013.10.001

Kuo, Y.-Y., Luo, J., & Brielmaier, J. (2015). Investigating Students' Use of Lecture Videos in Online Courses: A Case Study for Understanding Learning Behaviors via Data Mining. In F. Li, R. Klamma, M. Laanpere, J. Zhang, B. Manjón, & R. Lau (Eds.), Advances in Web-Based Learning - ICWL 2015. *Lecture Notes in Computer Science* (Vol. 9412, pp. 231-237). Springer. https://doi.org/10.1007/978-3-319-25515-6_21

Lan, A. S., Brinton, C. G., Yang, T. Y., & Chiang, M. (2017). Behavior-based latent variable model for learner engagement. In X. Hu, T. Barnes, A. Hershkovitz, & L. Paquette (Eds.), *Proceedings of the 10th International Conference on Educational Data Mining, EDM 2017* (pp. 64-71). EDM. http://educationaldatamining.org/EDM2017/proc_files/proceedings.pdf

Lau, K. H. V., Farooque, P., Leydon, G., Schwartz, M. L., Mark, R., & Moeller, J. J. (2018). Using learning analytics to evaluate a video-based lecture series. *Medical Teacher, 40*(1), 91-98. https://doi.org/10.1080/0142159X.2017.1395001

Lemay, D. J., & Doleck, T. (2020). Grade prediction of weekly assignments in MOOCS: mining video-viewing behavior. *Education and Information Technologies*, *25*(2), 1333-1342. https://doi.org/10.1007/s10639-019-10022-4

Li, M. F., Tang, X. P., Wu, W., & Liu, H. Bin. (2013). General models for estimating daily global solar radiation for different solar radiation zones in mainland China. *Energy Conversion and Management*, *70*, 139-148. https://doi.org/10.1016/j.enconman.2013.03.004

Li, Q., & Baker, R. (2018). The different relationships between engagement and outcomes across participant subgroups in Massive Open Online Courses. *Computers & Education*, *127*, 41-65. https://doi.org/10.1016/j.compedu.2018.08.005

Li, Z., & Chen, M. Y. (2019). Application of ANCOVA and MANCOVA in language assessment research. In V. Aryadoust & M. Raquel (Eds.), *Quantitative Data Analysis for Language Assessment Volume I: Fundamental Techniques* (pp. 198–218). Routledge. https://doi.org/10.4324/9781315187815-10

Lin, S.-Y., Aiken, J. M., Seaton, D. T., Douglas, S. S., Greco, E. F., Thoms, B. D., & Schatz, M. F. (2017). Exploring physics students' engagement with online instructional videos in an introductory mechanics course. *Physical Review Physics Education Research*, *13*(2), Article 020138. https://doi.org/10.1103/PhysRevPhysEducRes.13.020138

Lu, O., Huang, A., Huang, J., Lin, A., Ogata, H., & Yang, S. (2018). Applying Learning Analytics for the Early Prediction of Students' Academic Performance in Blended Learning. *Journal of Educational Technology & Society*, *21*(2), 220-232. https://doi.org/10.2307/26388400

Lu, Y., Zhang, S., Zhang, Z., Xiao, W., & Yu, S. (2017). A Framework for Learning Analytics Using Commodity Wearable Devices. *Sensors*, *17*(6), Article 1382. https://doi.org/10.3390/s17061382

Magdin, M., Cápay, M., & Mesárošová, M. (2011). Usage of interactive video in educational process to determine mental level and literacy of a learner. In *Proceedings - 14th International Conference on Interactive Collaborative Learning, ICL 2011* (pp. 510-513). IEEE. https://doi.org/10.1109/ICL.2011.6059637

Mangaroska, K., & Giannakos, M. (2018). Learning analytics for learning design: A systematic literature review of analytics-driven design to enhance learning. *IEEE Transactions on Learning Technologies*, *12*(4), 516-534. https://doi.org/10.1109/TLT.2018.2868673

Mar, C., Sohoni, S., & Craig, S. D. (2017). The effect of embedded questions in programming education. In *Proceedings - Frontiers in Education Conference, FIE, 2017*. IEEE. https://doi.org/10.1109/FIE.2017.8190494

Mayer, R. E. (2009). *Multimedia Learning* (2nd ed.). Cambridge University Press.

Mayer, R. E., & Moreno, R. (2003). Nine Ways to Reduce Cognitive Load in Multimedia Learning. *Educational Psychologist*, *38*(1), 43-52. https://doi.org/10.1207/S15326985EP3801_6

Mayer, R. E. (2014). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 43–71). Cambridge University Press. https://doi.org/10.1017/CBO9781139547369.005

Mbouzao, B., Desmarais, M. C., & Shrier, I. (2020). Early Prediction of Success in MOOC from Video Interaction Features. In I. Ibert Bittencourt, M. Cukurova, K. Muldner, R. Luckin, & E. Millán (Eds.), *Artificial Intelligence in Education* (Vol. 1, pp. 191-196). Springer. https://doi.org/10.1007/978-3-030-52240-7

McGowan, A., & Hanna, P. (2015). How video lecture capture affects student engagement in a higher education computer programming course: A study of attendance, video viewing behaviours and student attitude. In *EChallenges E-2015 Conference Proceedings*. IEEE. https://doi.org/10.1109/eCHALLENGES.2015.7440966

Mirriahi, N., & Vigentini, L. (2017). Analytics of Learner Video Use. In C. Lang, G. Siemens, A. Wise, & D. Gašević (Eds.), *Handbook of Learning Analytics* (pp. 251-267). SoLAR. https://doi.org/10.18608/hla17.022

Mongkhonvanit, K., Kanopka, K., & Lang, D. (2019). Deep knowledge tracing and engagement with MOOCs. In *LAK19: Proceedings of the 9th International Conference on Learning Analytics & Knowledge* (pp. 340-342). ACM. https://doi.org/10.1145/3303772.3303830

Mongy, S. (2007). A study on video viewing behavior: Application to movie trailer miner. *International Journal of Parallel, Emergent and Distributed Systems*, *22*(3), 163-172. https://doi.org/10.1080/17445760601125376

Morris, L. V., Finnegan, C., & Wu, S. S. (2005). Tracking student behavior, persistence, and achievement in online courses. *Internet and Higher Education*, *8*(3), 221-231. https://doi.org/10.1016/j.iheduc.2005.06.009

Mubarak, A. A. (2020). Predictive learning analytics using deep learning model in MOOCs' courses videos. *Education and Information Technologies*, *26*, 371-392. https://doi.org/10.1007/s10639-020-10273-6

Muñoz-García, A., & Villena-Martínez, M. D. (2021). Influences of learning approaches, student engagement, and satisfaction with learning on measures of sustainable behavior in a social sciences student sample. *Sustainability (Switzerland)*, *13*(2), Article 541. https://doi.org/10.3390/su13020541

Na, K. S., & Tasir, Z. (2017). A Systematic Review of Learning Analytics Intervention Contributing to Student Success in Online Learning. In *Proceedings - 5th International Conference on Learning and Teaching in Computing and Engineering, LaTiCE 2017* (pp. 62-68). IEEE. https://doi.org/10.1109/LaTiCE.2017.18

Nagy, J. T. (2018). Evaluation of online video usage and learning satisfaction: An extension of the technology acceptance model. *International Review of Research in Open and Distance Learning*, *19*(1), 160-185. https://doi.org/10.19173/irrodl.v19i1.2886

Nakayama, M., Yamamoto, H., & Santiago, R. (2007). Relationship between learner characteristics and learning performance in hybrid courses among Japanese students. In D. Remenyi (Ed.), *Proceedings of the International Conference on E-Learning, ICEL, 2007* (pp. 341-349). Academic Conferences.

Northrup, P. (2001). A Framework for Designing Interactivity into Web-Based Instruction. *Educational Technology*, *41*(2), 31-39.

Özhan, Ş. Ç., Aydın, F., Tepgeç, M., Aydın, M., & Yurdugül, H. (2019, October 30 - November 1). *Öğrenme Tasarımı İçin Öğrenme Analitiği Kullanımı: Video Analitik Örneği*. International Instructional Technologies and Teacher Education Symposium (ITTES 2019), Antalya, Turkey.

Park, S., & Yun, H. (2018). The Influence of Motivational Regulation Strategies on Online Students' Behavioral, Emotional, and Cognitive Engagement. *American Journal of Distance Education*, *32*(1), 43-56. https://doi.org/10.1080/08923647.2018.1412738

Pellas, N., & Kazanidis, I. (2015). On the value of Second Life for students' engagement in blended and online courses: A comparative study from the Higher Education in Greece. *Education and Information Technologies*, *20*(3), 445-466. https://doi.org/10.1007/s10639-013-9294-4

Peng, W. (2017). Research on model of student engagement in online learning. *Eurasia Journal of Mathematics, Science and Technology Education*, *13*(7), 2869-2882. https://doi.org/10.12973/eurasia.2017.00723a

Petan, S., Mocofan, M., & Vasiu, R. (2014). Enhancing Learning in Massive Open Online Courses through Interactive Video Conference. *The 9 Th International Scientific Conference ELearning and Software for Education*, *April*, 380–386.

Polier, S. (2019). *Forward-looking External Search as a Driver for Innovation*. Springer.

Puzziferro, M. (2008). Online technologies self-efficacy and self-regulated learning as predictors of final grade and satisfaction in college-level online courses. *International Journal of Phytoremediation*, *21*(1), 72-89. https://doi.org/10.1080/08923640802039024

Qiu, F., & Cui, Y. (2010). An analysis of user behavior in online video streaming. In *VLS-MCMR'10 - Proceedings of the 2010 ACM International Workshop on Very-Large-Scale Multimedia Corpus, Mining and Retrieval, Co-Located with ACM Multimedia 2010* (pp. 49-54). ACM. https://doi.org/10.1145/1878137.1878149

Quade, D. (1967). Rank analysis of covariance. *Journal of the American Statistical Association*, *62*(320), 1187–1200. https://doi.org/10.2307/2283769

Rajabalee, Y. B., & Santally, M. I. (2021). Learner satisfaction, engagement and performances in an online module: Implications for institutional e-learning policy. *Education and Information Technologies*, *26*(3), 2623-2656. https://doi.org/10.1007/s10639-020-10375-1

Redmond, P., Abawi, L. A., Brown, A., Henderson, R., & Heffernan, A. (2018). An online engagement framework for higher education. *Online Learning Journal*, *22*(1), 183-204. https://doi.org/10.24059/olj.v22i1.1175

Reich, J., & Ruipérez-Valiente, J. A. (2019). The MOOC pivot. *Science*, *363*(6423), 130-131. https://doi.org/10.1126/science.aav7958

Rice, P., Beeson, P., & Blackmore-Wright, J. (2019). Evaluating the Impact of a Quiz Question within an Educational Video. *TechTrends*, *63*(5), 522-532. https://doi.org/10.1007/s11528-019-00374-6

Rienties, B., Boroowa, A., Cross, S., Kubiak, C., & Mayles, K. (2016). Analytics4Action Evaluation Framework: A Review of Evidence-Based Learning Analytics Interventions at the Open University UK. *Journal of Interactive Media in Education*, *2016*(1), 1-11. https://doi.org/10.5334/jime.394

Rienties, B., Cross, S., & Zdrahal, Z. (2016). Implementing a Learning Analytics Intervention and Evaluation Framework: What Works? In B. Kei Daniel (Ed.), *Big Data and Learning Analytics in Higher Education: Current Theory and Practice* (pp. 147-166). https://doi.org/10.1007/978-3-319-06520-5_10

Rienties, B., & Toetenel, L. (2015). "Scaling up" learning design: impact of learning design activities on LMS behavior and performance. In *LAK '15: Proceedings of the Fifth International Conference on Learning Analytics and Knowledge* (pp 315-319). ACM. https://doi.org/10.1145/2723576.2723600

Risko, E. F., Foulsham, T., Dawson, S., & Kingstone, A. (2013). The Collaborative Lecture Annotation System (CLAS): A New TOOL for Distributed Learning. *IEEE Transactions on Learning Technologies* *6*(1), 4-13. https://doi.org/10.1109/TLT.2012.15

Roach, V., & Lemasters, L. (2006). Satisfaction with online learning: A comparative descriptive study. *Journal of Interactive Online Learning*, *5*(3), 317-332. http://www.ncolr.org/issues/jiol/v5/n3/satisfaction-with-online-learning-a-comparative-descriptive-study.html

Ronchetti, M. (2013). Videolectures ingredients that can make analytics effective. In M. N. Giannakos, K. Chorianopoulos, M. Ronchetti, P. Szegedi, & S. Teasley (Eds.), *Proceedings of the LAK 2013 Workshop on Analytics on Video-based Learning* (Vol. 983, pp. 15-20. CEUR. http://ceur-ws.org/Vol-983/paper4.pdf

Şahin, M. (2018). *E-Öğrenme Ortamlarına Yönelik Öğrenme Analitiklerine Dayalı Bir Müdahale Motoru Tasarımı* [Doctoral dissertation, Hacettepe University, Ankara, Turkey. http://www.openaccess.hacettepe.edu.tr:8080/xmlui/handle/11655/5312

Şahin, M., & Yurdugül, H. (2019). An intervention engine design and development based on learning analytics: the intelligent intervention system (In2S). *Smart Learning Environments*, *6*(1), Article 18. https://doi.org/10.1186/s40561-019-0100-7

Scheffel, M. (2017). *The Evaluation Framework for Learning Analytics*. https://core.ac.uk/download/pdf/95759389.pdf

Schielzeth, H., Dingemanse, N. J., Nakagawa, S., Westneat, D. F., Allegue, H., Teplitsky, C., Réale, D., Dochtermann, N. A., Garamszegi, L. Z., & Araya-Ajoy, Y. G. (2020). Robustness of linear mixed-effects models to violations of distributional assumptions. *Methods in Ecology and Evolution*, *11*(9), 1141-1152. https://doi.org/10.1111/2041-210X.13434

Shi, C., Fu, S., Chen, Q., & Qu, H. (2015). VisMOOC: Visualizing video clickstream data from Massive Open Online Courses. In S. Liu, G. Scheuermann, & S. Takahashi (Eds.), *Proceedings - IEEE Pacific Visualization Symposium*, *2015* (pp. 159-166). IEEE. https://doi.org/10.1109/PACIFICVIS.2015.7156373

Shukor, N. A., Tasir, Z., Van der Meijden, H., & Harun, J. (2014). A Predictive Model to Evaluate Students' Cognitive Engagement in Online Learning. *Procedia - Social and Behavioral Sciences*, *116*, 4844-4853. https://doi.org/10.1016/j.sbspro.2014.01.1036

Siemens, G. (2013). Learning Analytics : The Emergence of a Discipline. *American Behavioral Scientist*, *57*(10), 1380-1400. https://doi.org/10.1177/0002764213498851

Sinha, T., Jermann, P., Li, N., & Dillenbourg, P. (2014). *Your click decides your fate: Inferring Information Processing and Attrition Behavior from MOOC Video Clickstream Interactions*. http://arxiv.org/abs/1407.7131

Smith, M. C., & Winking-diaz, A. (2004). Increasing Students' Interactivity in an Online Course M Cecil Smith and Amy Winking-Diaz. *The Journal of Interactive Online Learning*, *2*(3). http://www.ncolr.org/issues/jiol/v2/n3/increasing-students-interactivity-in-an-online-course.html

Soni, A., Kumar, V., Kaur, R., & Hemavathi, D. (2018). *Predicting Student Performance*. *International Journal of Pure and Applied Mathematics*, *119*(12/SI), 221-227. https://acadpubl.eu/hub/2018-119-12/articles/7/1591.pdf

Sousa, F., & Martins, N. (2021). Learning Experience Design: Instructional Design Applied to the Onboarding of Digital Products Filipa. In N. Martins, & D. Brandão (Eds.), *Advances in Design and Digital Communication* (pp. 45-57). Springer. https://doi.org/10.1007/978-3-030-61671-7_5

Strachota, E. M. (2003). *Student Satisfaction in Online Courses : An Analysis of The Impact of Learner-Content, and Learner-Technology Interaction* [Doctoral dissertation, University of Wisconsin-Milwaukee]. https://www.proquest.com/openview/e70d11e01919cc387d0135a0fa7d2c10/1?pq-origsite=gscholar&cbl=18750&diss=y

Sullivan, R. A., Kuzel, A. M. H., Vaandering, M. E., & Chen, W. (2017). The Association of Physical Activity and Academic Behavior: A Systematic Review. *Journal of School Health*, *87*(5), 388-398. https://doi.org/10.1111/josh.12502

Toetenel, L., & Rienties, B. (2016). Analysing 157 learning designs using learning analytic approaches as a means to evaluate the impact of pedagogical decision making. *British Journal of Educational Technology*, *47*(5), 981-992. https://doi.org/10.1111/bjet.12423

Ullrich, C., Shen, R., & Xie, W. (2013). Analyzing student viewing patterns in lecture videos. In *Proceedings - 2013 IEEE 13th International Conference on Advanced Learning Technologies, ICALT 2013* (pp. 115-117). IEEE. https://doi.org/10.1109/ICALT.2013.38

Van der Meij, H., & Böckmann, L. (2020). Effects of embedded questions in recorded lectures. *Journal of Computing in Higher Education*, *33*, 235-254. https://doi.org/10.1007/s12528-020-09263-x

Van Der Sluis, F., Ginn, J., & Van Der Zee, T. (2016). Explaining student behavior at scale: The influence of video complexity on student dwelling time. *L@S 2016 - Proceedings of the 3rd 2016 ACM Conference on Learning at Scale* (pp. 51-60). ACM. https://doi.org/10.1145/2876034.2876051

Visser, L. N. C., Hillen, M. A., Verdam, M. G. E., Bol, N., de Haes, H. C. J. M., & Smets, E. M. A. (2016). Assessing engagement while viewing video vignettes; validation of the Video Engagement Scale (VES). *Patient Education and Counseling*, *99*(2), 227-235. https://doi.org/10.1016/j.pec.2015.08.029

Volungevičienė, A., Duart, J. M., Naujokaitienė, J., Tamoliūnė, G., & Misiulienė, R. (2019). Learning analytics: Learning to think and make decisions. *Journal of Educators Online*, *16*(2). https://doi.org/10.9743/jeo.2019.12.2.13

Vural, O. F. (2013). The Impact of a Question-Embedded Video-Based Learning Tool on E-Learning, Educational Sciences: Theory and Practice. *Educational Sciences: Theory & Practice*, *13*(2), 1315-1323.

Wachtler, J., Hubmann, M., Zöhrer, H., & Ebner, M. (2016). An analysis of the use and effect of questions in interactive learning-videos. *Smart Learning Environments*, *3*(1), Article 13. https://doi.org/10.1186/s40561-016-0033-3

Wachtler, J., Khalil, M., Taraghi, B., & Ebner, M. (2016). On using learning analytics to track the activity of interactive MOOC videos. In M. N. Giannakos, D. G. Sampson, Ł. Kidziński, & A. Pardo (Eds.), *Proceedings of the LAK 2016 Workshop on Smart Environments and Analytics in Video-Based Learning* (Vol. 1579, pp. 8-17). CEUR.

Wang, M.-T., & Eccles, J. S. (2012). Adolescent Behavioral, Emotional, and Cognitive Engagement Trajectories in School and Their Differential Relations to Educational Success. *Journal of Research on Adolescence*, *22*(1), 31-39. https://doi.org/10.1111/j.1532-7795.2011.00753.x

West, D., Heath, D., & Huijser, H. (2016). Let's talk learning analytics: A framework for implementation in relation to student retention. *Journal of Asynchronous Learning Network*, *20*(2). https://doi.org/10.24059/olj.v20i2.792

Wong, B. T. M. (2017). Learning analytics in higher education: an analysis of case studies. *Asian Association of Open Universities Journal*, *12*(1), 21-40. https://doi.org/10.1108/aaouj-01-2017-0009

Wong, B. T.-M, & Li, K. C. (2020). A review of learning analytics intervention in higher education (2011–2018). *Journal of Computers in Education*, *7*(1), 7-28. https://doi.org/10.1007/s40692-019-00143-7

Xie, S., Yuen, W.-K., Huang, L., & Ng, V. T. Y. (2017). Student engagement with video-watching and flipped class behaviors. In *Proceedings of the International Conference on E-Learning (ICEMEET 2016)* (pp. 163-170). Springer. https://doi.org/10.2991/icemeet-16.2017.54

Yang, S. J. H., Lu, O. H. T., Huang, A. Y. Q., Huang, J. C. H., Ogata, H., & Lin, A. J. Q. (2018). Predicting Students' Academic Performance Using Multiple Linear Regression and Principal Component Analysis. *Journal of Information Processing*, *26*, 170-176. https://doi.org/10.2197/ipsjjip.26.170

Yang, T.-Y., Brinton, C. G., Joe-Wong, C., & Chiang, M. (2017). Behavior-Based Grade Prediction for MOOCs Via Time Series Neural Networks. *IEEE Journal on Selected Topics in Signal Processing*, *11*(5), 716-728. https://doi.org/10.1109/JSTSP.2017.2700227

Yeboah, A. K., & Smith, P. (2016). Relationships between minority students online learning experiences and academic performance. *Online Learning*, *20*(4). https://doi.org/10.24059/olj.v20i4.577

Yousef, A. M. F., Chatti, M. A., & Schroeder, U. (2014). Video-based learning :A critical analysis of the research published in 2003-2013 and future visions. In S. White (Ed.), *ELmL 2014: The Sixth International Conference on Mobile, Hybrid and On-Line Learning* (pp. 112-119). IARIA.

Yu, C.-H., Wu, J., & Liu, A.-C. (2019). Predicting Learning Outcomes with MOOC Clickstreams. *Education Sciences, 9*(2), Article 104. https://doi.org/10.3390/educsci9020104

Yürüm, O. R., & Yıldırım, S. (2020). Video Analitikler. In T. Güyer, H. Yurdugül, & S. Yıldırım (Eds.), *Eğitsel Veri Madenciliği ve Öğrenme Analitikleri* (pp. 185-202). Anı Yayıncılık.

Zhang, D. (2005). Interactive Multimedia-Based E-Learning: A Study of Effectiveness. *International Journal of Phytoremediation*, *21*(1), 149-162. https://doi.org/10.1207/s15389286ajde1903_3

Zhang, D., Zhou, L., Briggs, R. O., & Nunamaker, J. F. (2006). Instructional video in e-learning: Assessing the impact of interactive video on learning effectiveness. *Information and Management*, *43*(1), 15-27. https://doi.org/10.1016/j.im.2005.01.004

Zhang, J.-H., Zou, L.-Cong, Miao, J.-J., Zhang, Y.-X., Hwang, G.-J., & Zhu, Y. (2020). An individualized intervention approach to improving university students' learning performance and interactive behaviors in a blended learning environment. *Interactive Learning Environments*, *28*(2), 231-245. https://doi.org/10.1080/10494820.2019.1636078

Zhang, Y., & Wildemuth, B. M. (2005). Qualitative Analysis of Content. *Human Brain Mapping*, *30*(7), 2197-2206.

Zhou, C., Guo, Y., Chen, Y., Nie, X., & Zhu, W. (2014). Characterizing user watching behavior and video quality in mobile devices. In *Proceedings - International Conference on Computer Communications and Networks, ICCCN*. IEEE. https://doi.org/10.1109/ICCCN.2014.6911828

# APPENDICES

## APPENDIX A

## BOX PLOTS FOR VIDEO LECTURES

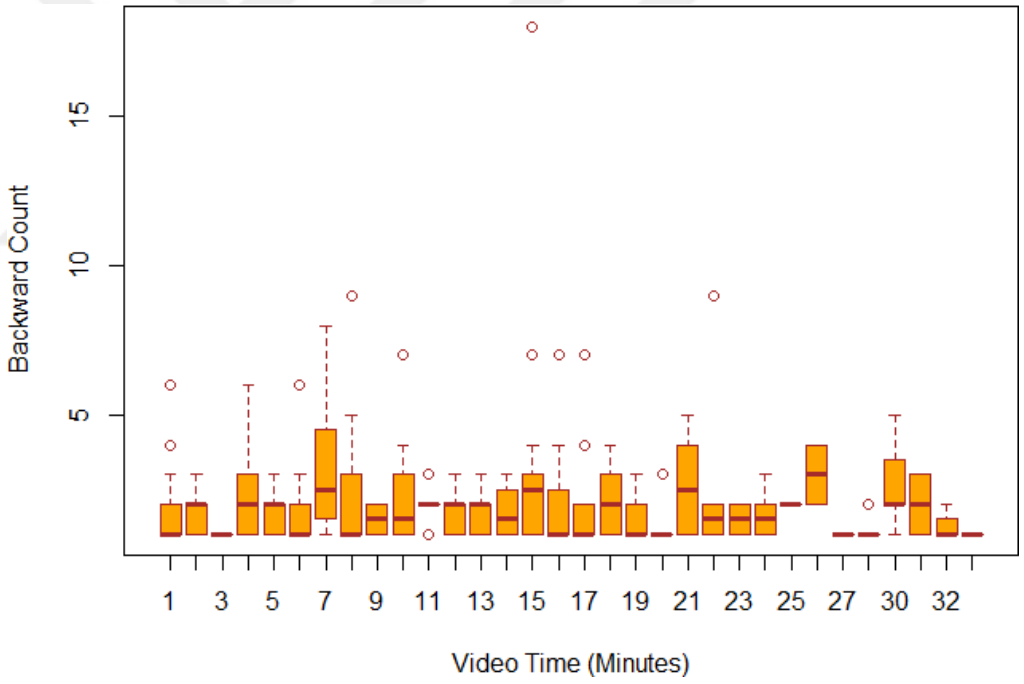



**Figure A1. Box Plot for Backward Count per Minute**
**(obtained from 27 students watching video lecture about Skinner; 32:40 minutes)**

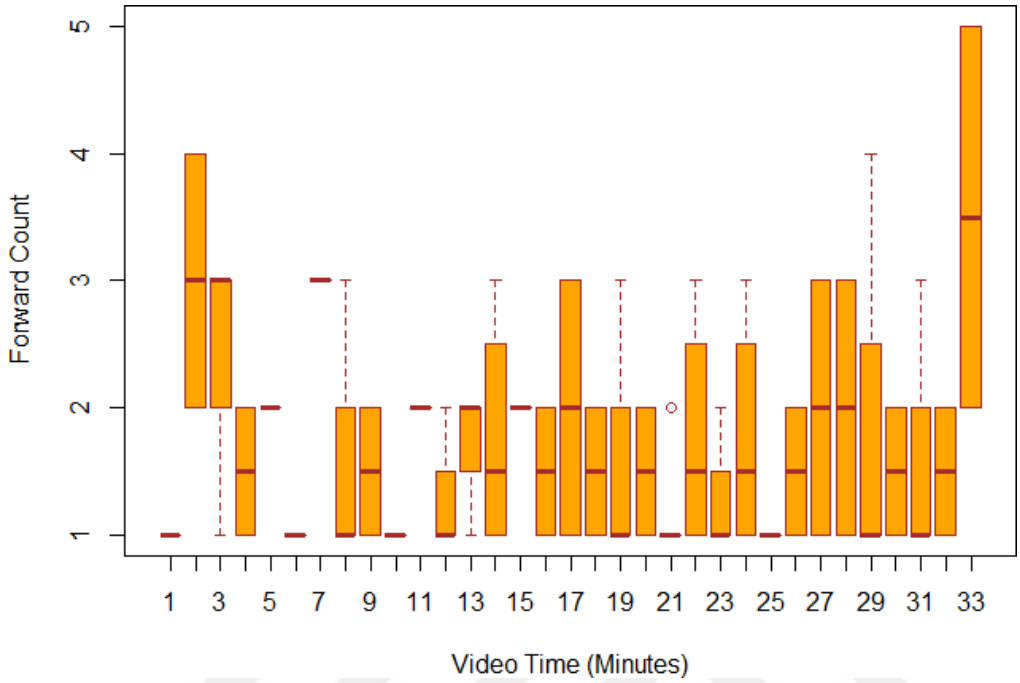



**Figure A2. Box Plot for Forward Count per Minute**
**(obtained from 27 Students watching video lecture about Skinner; 32:40 minutes)**

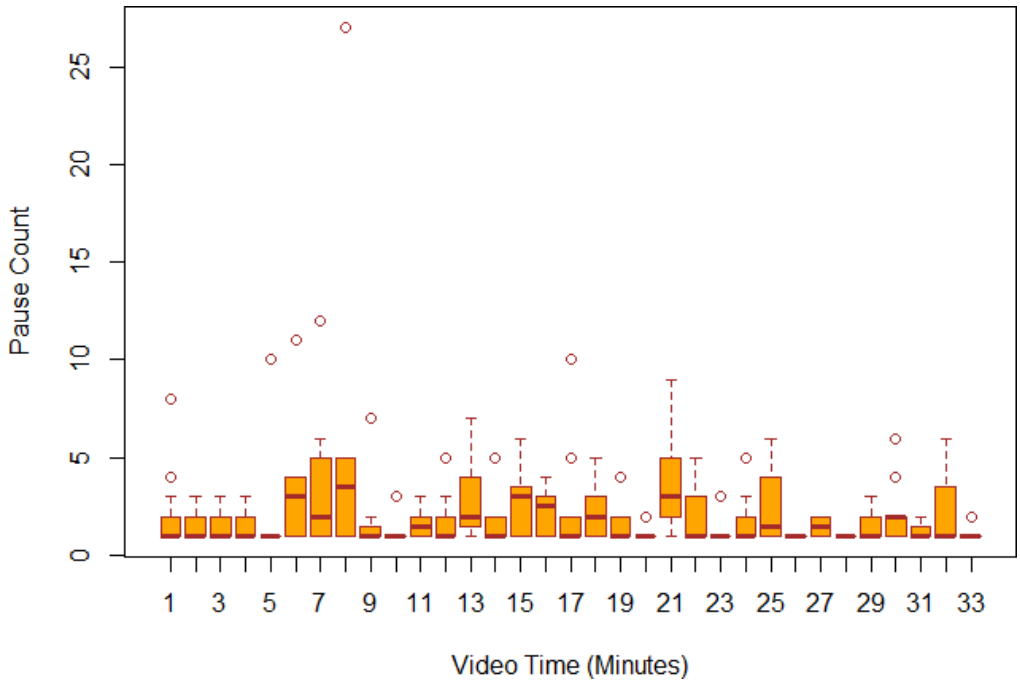



**Figure A3. Box Plot for Pause Count per Minute**
**(obtained from 27 students watching video lecture about Skinner; 32:40 minutes)**

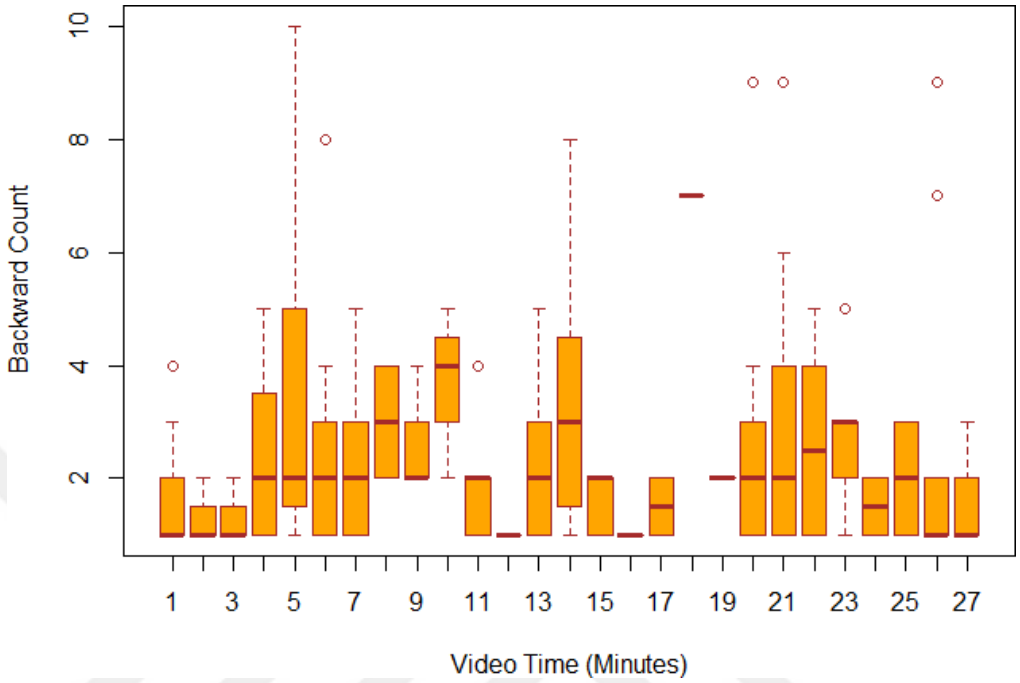



**Figure A4. Box Plot for Backward Count per Minute**
**(obtained from 25 students watching video lecture about Piaget; 27:06 minutes)**

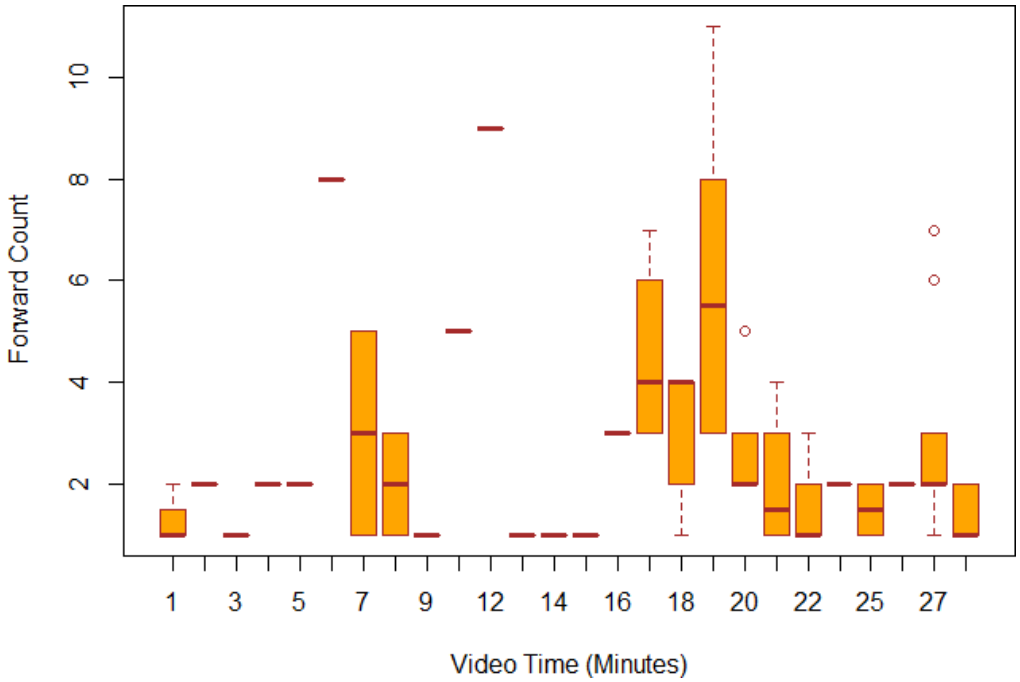



**Figure A5. Box Plot for Forward Count per Minute**
**(obtained from 25 students watching video lecture about Piaget; 27:06 minutes)**

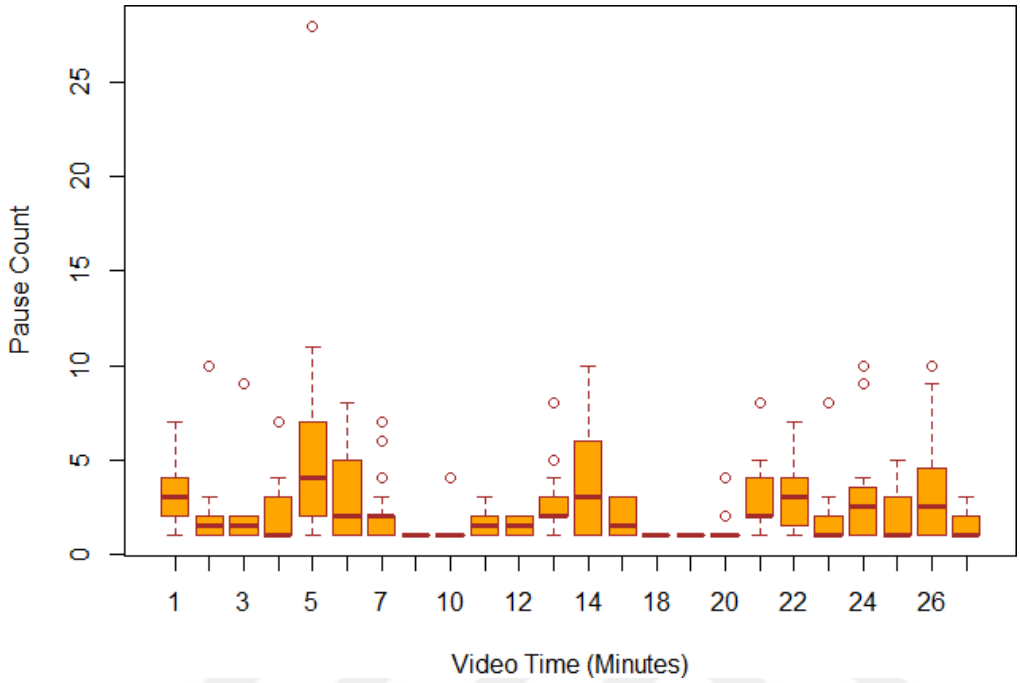



**Figure A6. Box Plot for Pause Count per Minute**
**(obtained from 25 students watching video lecture about Piaget; 27:06 minutes)**

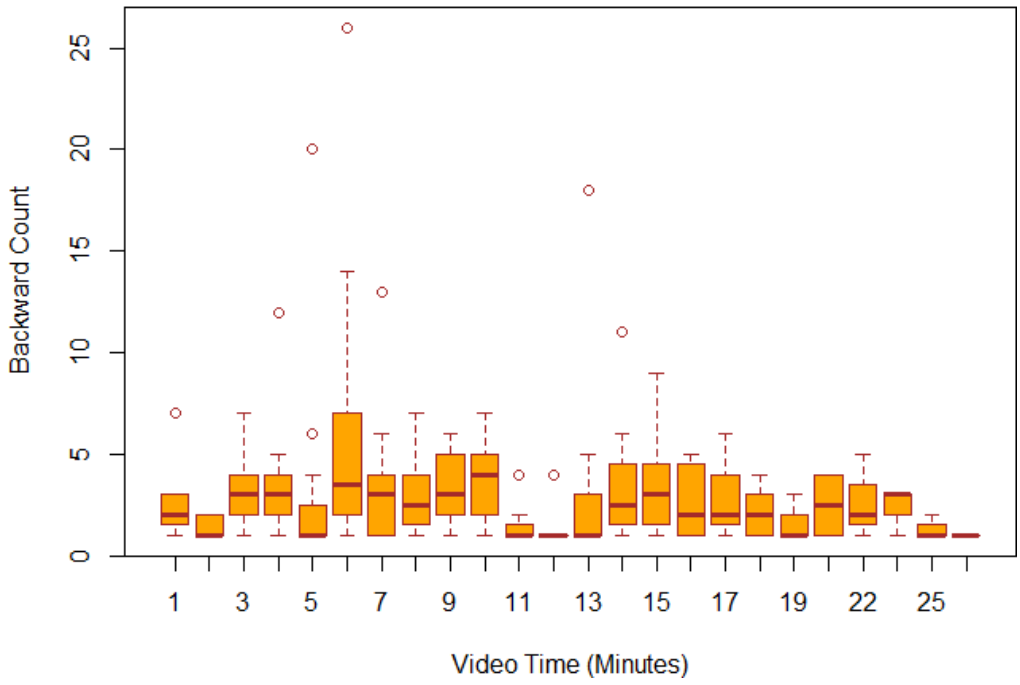



**Figure A7. Box Plot for Backward Count per Minute**
**(obtained from 25 students watching video lecture about Vygotsky; 26:13 minutes)**

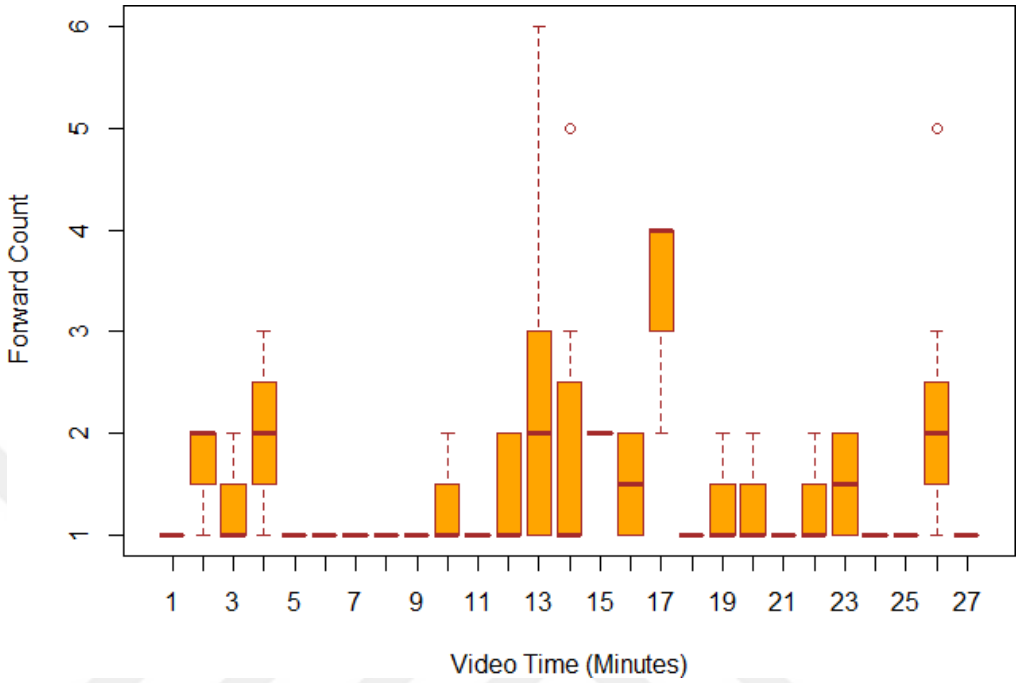



**Figure A8. Box Plot for Forward Count per Minute**
**(obtained from 25 students watching video lecture about Vygotsky; 26:13 minutes)**

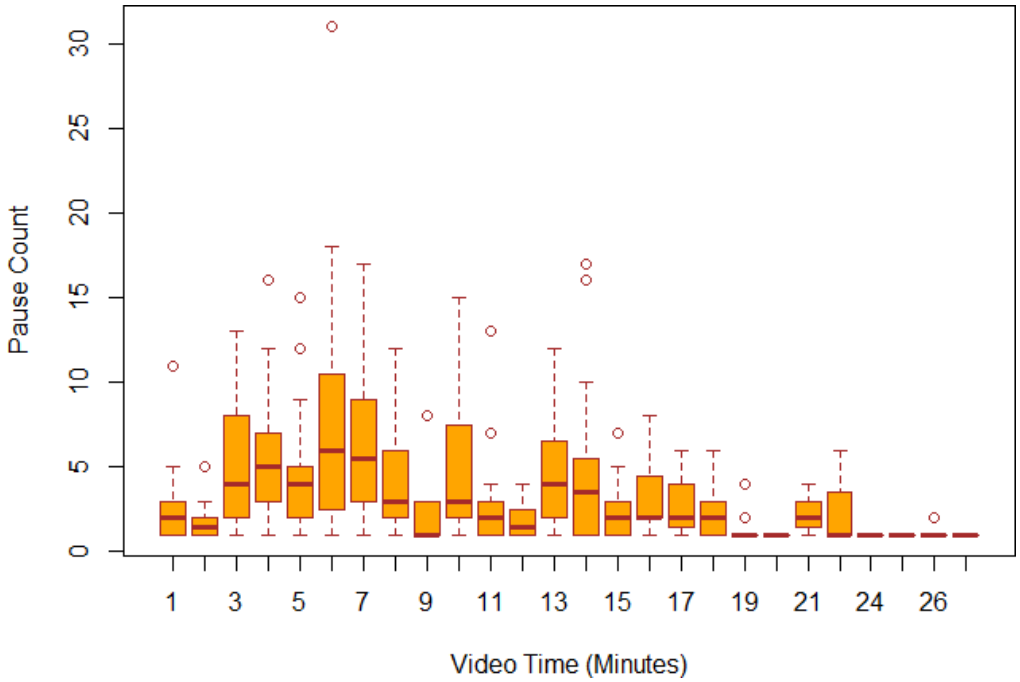
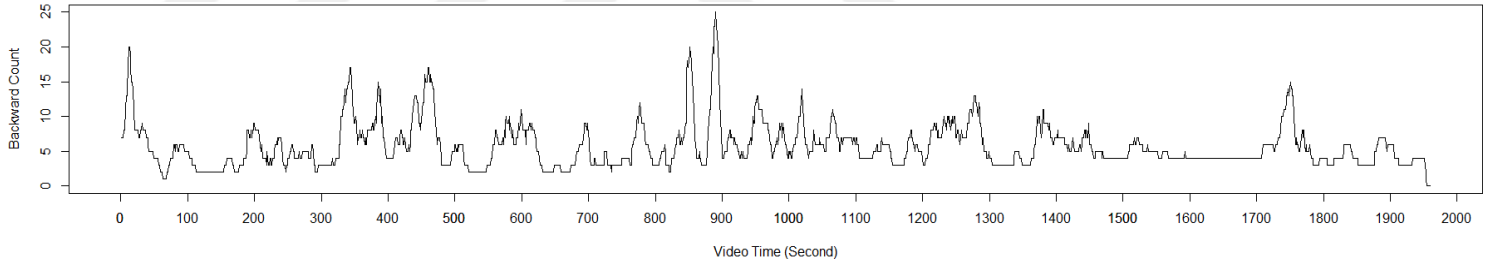



**Figure A9. Box Plot for Pause Count per Minute**
**(obtained from 25 students watching video lecture about Vygotsky; 26:13 minutes)**

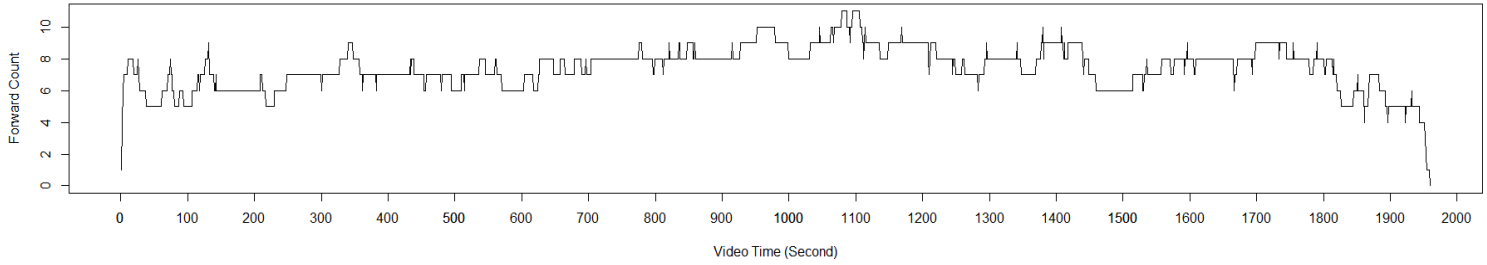
# APPENDIX B

## LINE GRAPHS FOR VIDEO LECTURE 1

a)




b)




c)

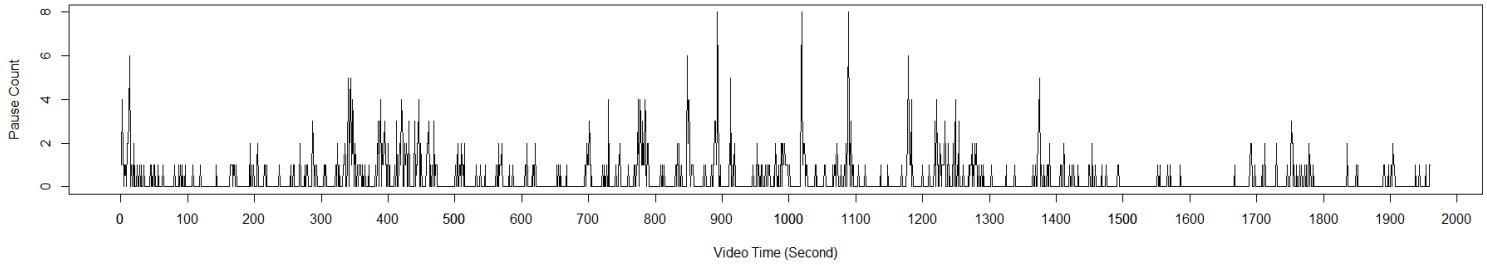
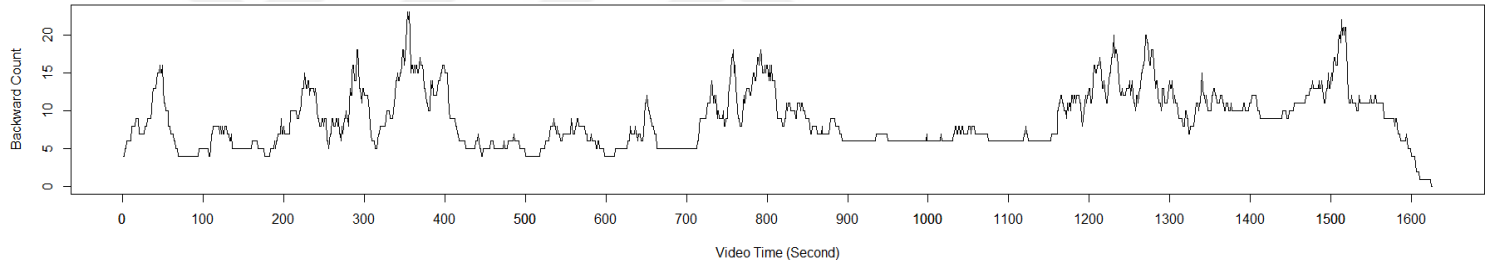


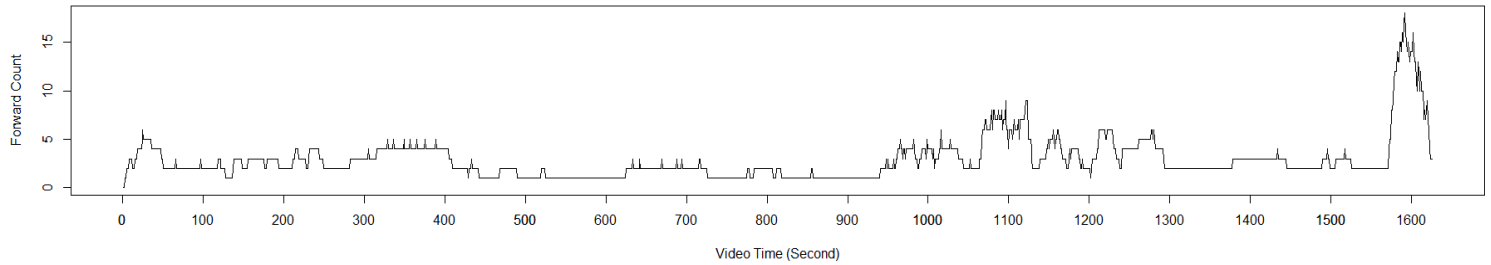

**Figure B1. Line Graphs for Backward, Forward and Pause Counts per Second**
**(obtained from 27 students watching video lecture about Skinner; 32:40 minutes)**

## LINE GRAPHS FOR VIDEO LECTURE 2

**a)**




**b)**




**c)**

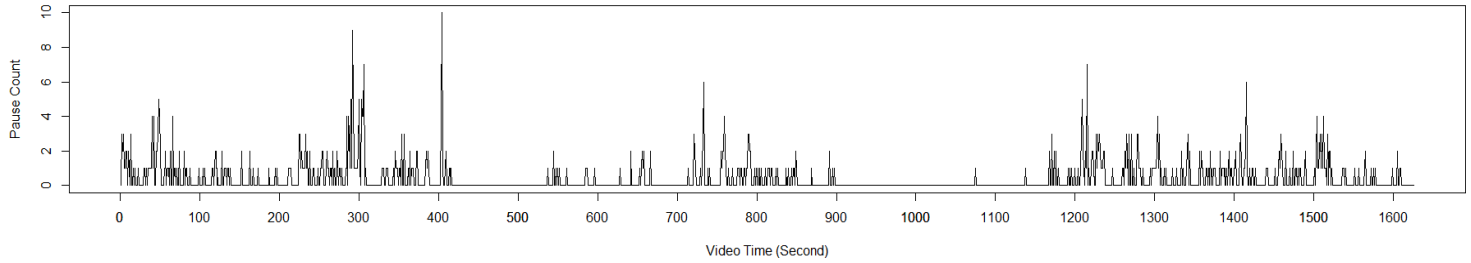
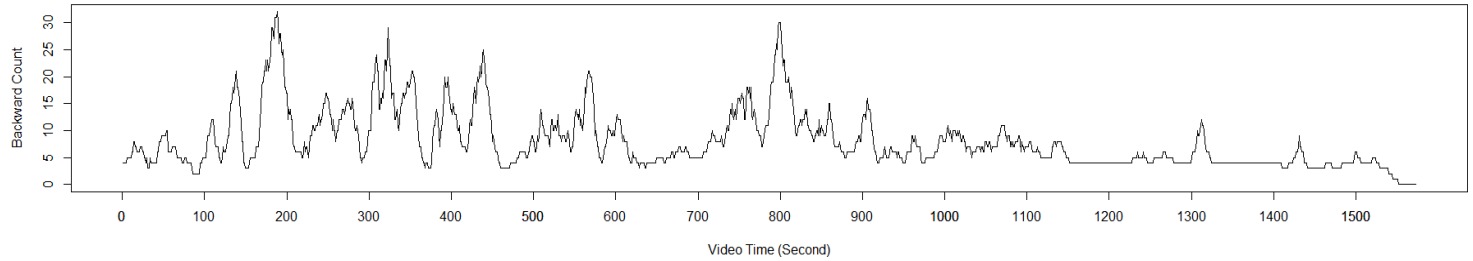


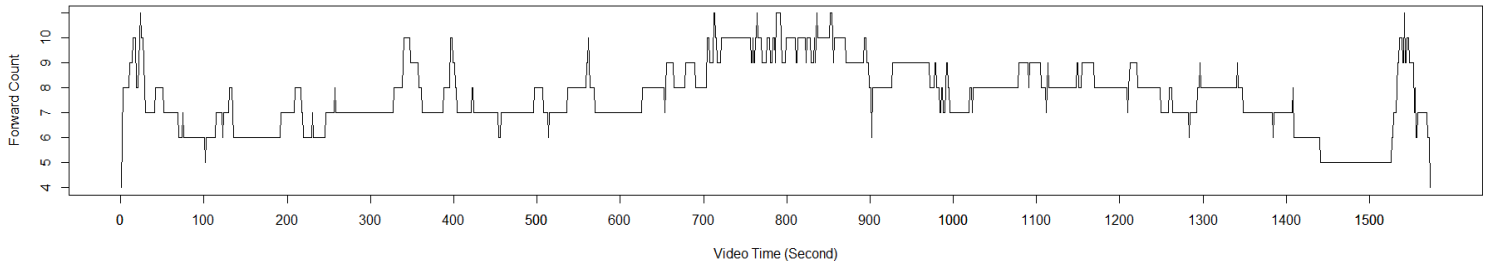

**Figure B2. Line Graphs for Backward, Forward and Pause Counts per Second**
**(obtained from 25 students watching video lecture about Piaget; 27:06 minutes)**

## LINE GRAPHS FOR VIDEO LECTURE 3

**a)**




**b)**




**c)**

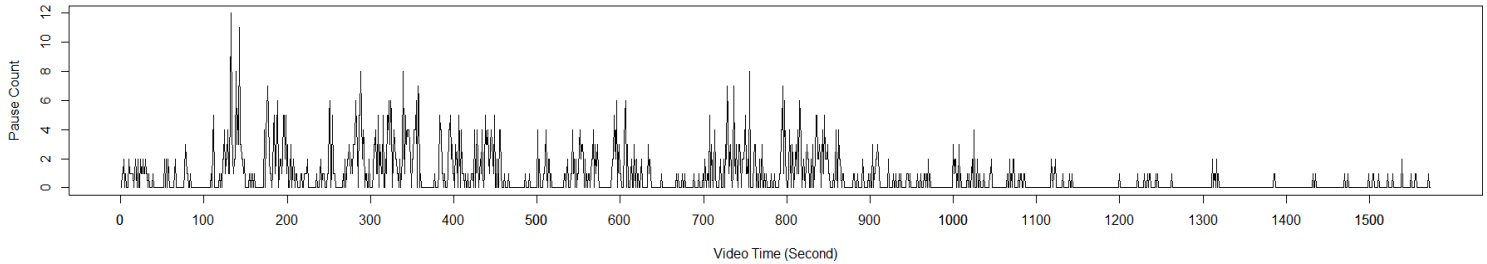



**Figure B3. Line Graphs for Backward, Forward and Pause Counts per Second**
**(obtained from 25 students watching video lecture about Vygotsky; 26:13 minutes)**

# APPENDIX C

## EXAMPLE WATCHING BEHAVIOR GRAPHS SHOWN TO STUDENTS DURING INTERVIEW

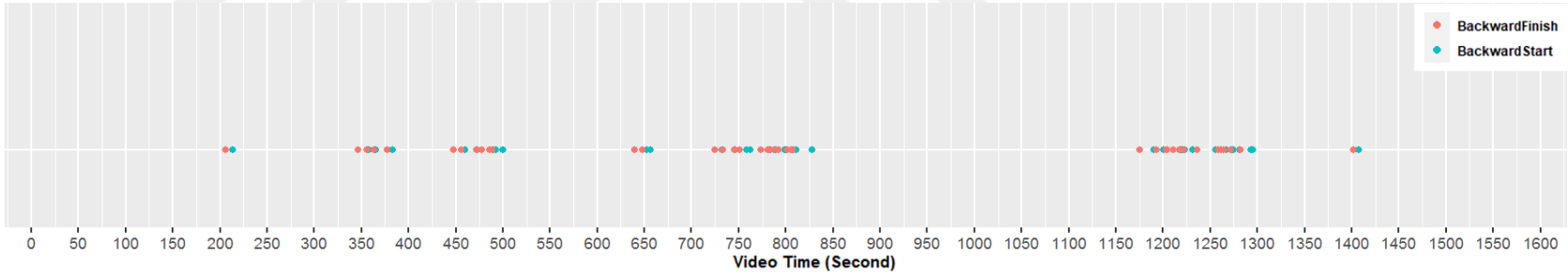



**Figure C1. Example of User's Backward Graph**

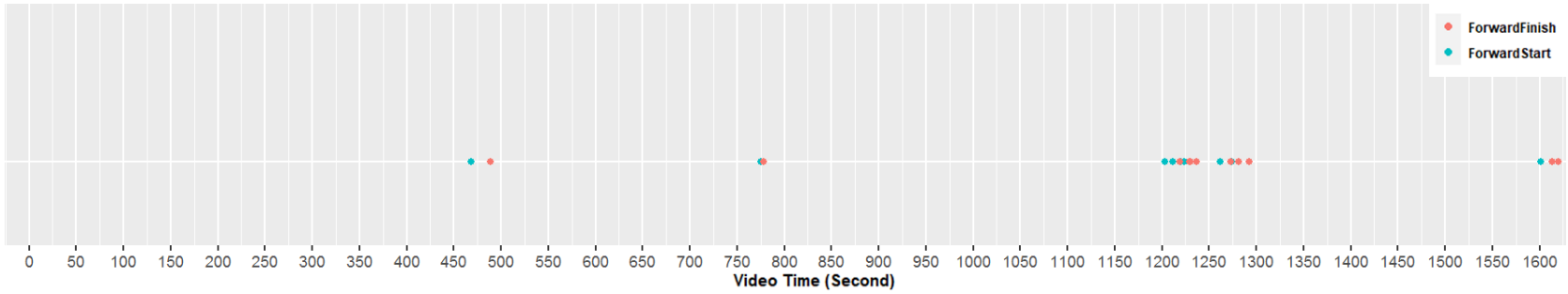



**Figure C2. Example of User's Pause Graph**

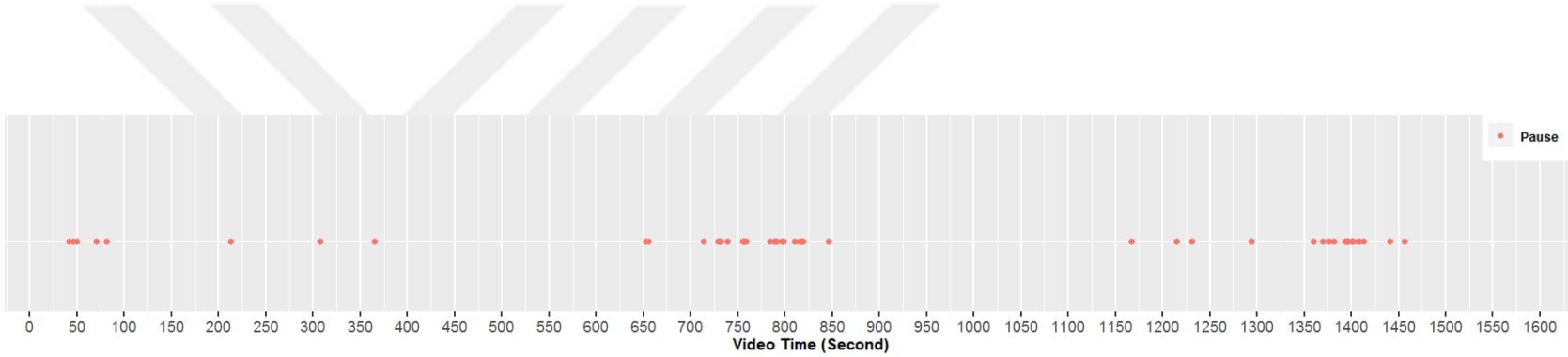



**Figure C3. Example of User's Forward Graph**

# APPENDIX D

# ETHICAL FORM

UYGULAMALI ETİK ARAŞTIRMA MERKEZİ
APPLIED ETHICS RESEARCH CENTER

ORTA DOĞU TEKNİK ÜNİVERSİTESİ
MIDDLE EAST TECHNICAL UNIVERSITY

DUMLUPINAR BULVARI 06800
ÇANKAYA ANKARA/TURKEY
T: +90 312 210 22 91
F: +90 312 210 79 59
ueam@metu.edu.tr
www.ueam.metu.edu.tr

Sayı: 28620816 / 202

09 Nisan 2019

Konu: Değerlendirme Sonucu

Gönderen: ODTÜ İnsan Araştırmaları Etik Kurulu (İAEK)

İlgi: İnsan Araştırmaları Etik Kurulu Başvurusu

**Sayın Prof. Dr. Soner YILDIRIM**

Danışmanlığını yaptığınız **Ozan Raşit YÜRÜM'ün "Öğrencilerin Video İzleme Davranışlarının İncelenmesi ve Video-tabanlı Öğrenme Analitiği Çerçevesi Geliştirilmesi"** başlıklı araştırması İnsan Araştırmaları Etik Kurulu tarafından uygun görülmüş ve **189-ODTÜ-2019** protokol numarası ile onaylanmıştır.

Saygılarımızla bilgilerinize sunarız

## APPENDIX E

## SCALES

### The Situational Motivation Scale (Guay et al., 2000)

Why are you currently engaged in this activity?

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 1. Because I think that this activity is interesting | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 2. Because I am doing it for my own good | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 3. Because I am supposed to do it | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 4. There may be good reasons to do this activity, but personally I don't see any | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 5. Because I think that this activity is pleasant | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 6. Because I think that this activity is good for me | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 7. Because it is something that I have to do | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 8. I do this activity but I am not sure if it is worth it | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 9. Because this activity is fun | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 10. By personal decision | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 11. Because I don't have any choice | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 12. I don't know; I don't see what this activity brings me | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 13. Because I feel good when doing this activity | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 14. Because I believe that this activity is important for me | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 15. Because I feel that I have to do it | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 16. I do this activity, but I am not sure it is a good thing to pursue it | 1 | 2 | 3 | 4 | 5 | 6 | 7 |

Codification key: Intrinsic motivation: Items 1, 5, 9, 13; Identified regulation: Items 2, 6, 10, 14; External regulation: Items 3,7, 11, 15; Amotivation: Items 4, 8, 12, 16.

### Learner Satisfaction Scale (Donkor, 2011)

| Element of satisfaction with the video-based instructional materials (VBIMs) |
|---|
| I find the video lessons enjoyable. |
| The VBIMs have contributed greatly to my acquisition of relevant skills in Block-Laying and Concreting. |
| I find the video lessons to be effective in meeting the learning objectives. |
| I would describe the video lessons as being highly interesting. |
| I would recommend use of the video lessons to my colleagues. |
| The video lessons make me spend more time studying to acquire practical skills. |
| I am satisfied with my learning from the VBIMs. |
| Resultant mean for all elements of satisfaction with the video-based instructional materials |

# Video Engagement Scale (Visser et al., 2016)

## VES® Video Engagement Scale

® L. Visser, M. Hillen, M. Verdam, N. Bol, H. de Haes, E. Smets

**Instructions:**

With the following statements we would like to get an idea of how **engaged** you were with the video while you were viewing it.

Please indicate how much you agree with the following statements. There are no right or wrong answers. You can place an 'X' in the box that represents your opinion the best.

If you completely **disagree** with the statement, you can place an 'X' in the extreme **left box**. The more you **agree** with the statement, the more you can place an 'X' in the direction of the extreme **right box**.

| | Completely disagree | | | | | | Completely agree |
|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 1. During viewing I was fully concentrated on the video | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 2. During viewing it was as if I was present at the events depicted in the video | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 3. When I was viewing the video, my thoughts were only with the video | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 4. After the video was finished, I had the feeling I came back into the 'real' world | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 5. When I had been viewing for a while, it seemed as if I had become the [video character]* in my thoughts | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 6. I empathized with the [video character]* | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 7. The video affected me | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 8. When I was viewing the video, I was in the world of the video in my thoughts | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 9. During viewing, I felt [particular emotion]* when the [video character]* felt [particular emotion]* | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 10. I found the video moving | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 11. I felt for the [video character]* | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 12. During viewing, I was hardly aware of the space around me | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 13. I had the feeling I went through what the [video character]* went through | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 14. In my imagination, it was as if I were the [video character]* | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| 15. Because of the video, feelings arose in me | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |

*Notes.* The VES total score can be calculated by a adding up scores on the 15 individual items.
*To be specified depending on the video content

## APPENDIX F

### Click Reasons Questionnaire (Created from Framework)

| **During the video lecture,** | 1 (Never) | 2 (Sometimes) | 3 (Often) | 4 (Usually) | 5 (Always) |
|---|---|---|---|---|---|
| I took notes. | | | | | |
| I lost concentration. | | | | | |
| I did not understand the content. | | | | | |
| I faced language issues. | | | | | |
| I went back to important points. | | | | | |
| I skipped unimportant points. | | | | | |
| I wanted to finish quickly. | | | | | |

## APPENDIX G

# EXAMPLE INTERVENTIONS




**Figure G1. Intervention 1**

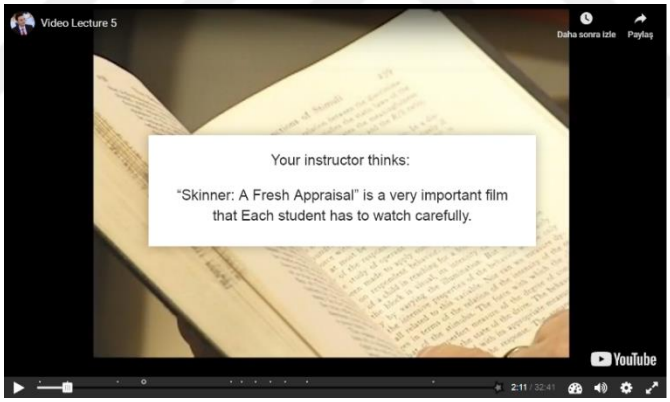



**Figure G2. Intervention 2**

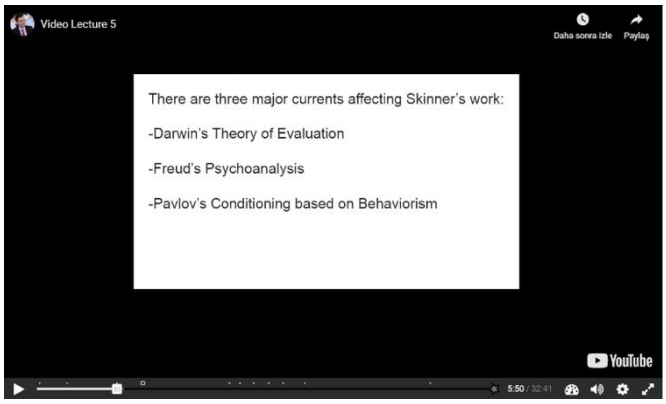



**Figure G3. Intervention 3**

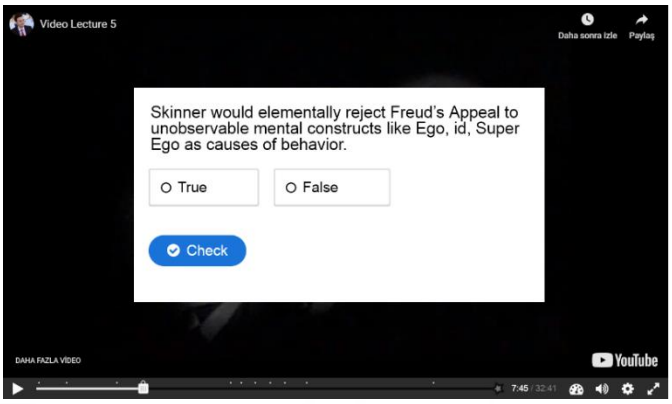



**Figure G4. Intervention 4**




**Figure G5. Intervention 5**



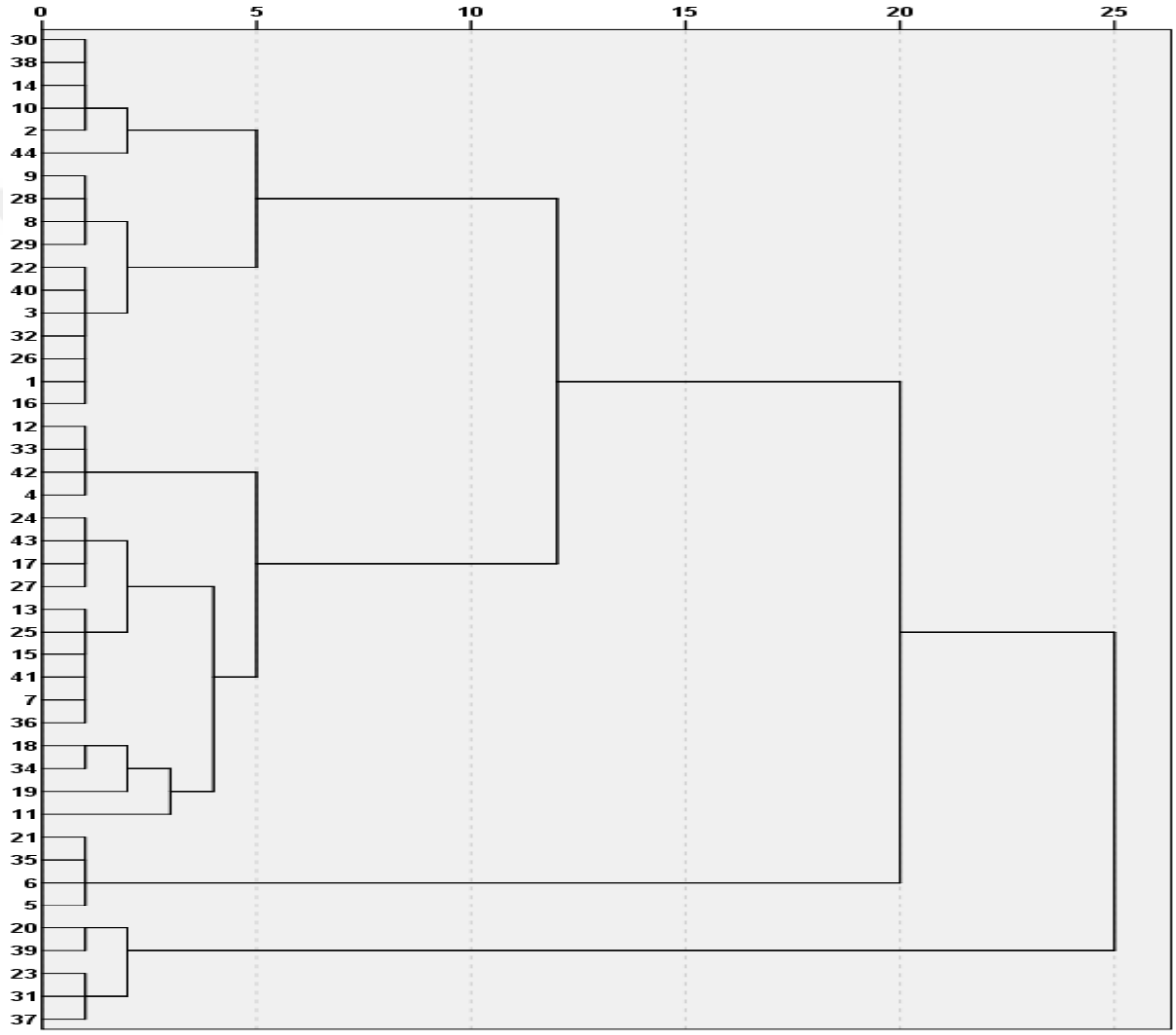

**Figure G6. Intervention 6**




**Figure G7. Intervention 7**

## APPENDIX H

## DENDROGRAM USING AVERAGE LINKAGE BETWEEN GROUPS FOR HIERARCHICAL CLUSTERING




* Clustering based on percentage of pauses, forwards, backwards was performed by analyzing the dendrogram (see Section 4.2.2.2).

# APPENDIX I

## CROSS VALIDATION RESULTS FOR LEARNING PERFORMANCE PREDICTION

**Table Ia. Random Split (Stage 1 RMSE Values)**

| Split | XGBoost | Robust LMM | Baseline |
|---|---|---|---|
| 1 | 14.76 | 14.70 | 18.10 |
| 2 | 13.83 | 9.98 | 11.41 |
| 3 | 16.40 | 20.46 | 24.81 |
| 4 | 16.49 | 15.03 | 18.58 |
| 5 | 11.95 | 14.12 | 15.75 |
| **Mean** | **14.69** | 14.86 | 17.73 |

**Table Ib. One Person Leave Out (Stage 1 RMSE Values)**

| Person | XGBoost | Robust LMM | Baseline |
|---|---|---|---|
| 1 | 10.05 | 11.12 | 12.33 |
| 2 | 17.01 | 24.72 | 27.38 |
| 3 | 29.64 | 19.17 | 11.13 |
| 4 | 4.20 | 6.36 | 16.66 |
| 5 | 8.56 | 7.21 | 16.85 |
| 6 | 0.36 | 3.08 | 15.54 |
| 7 | 23.72 | 24.70 | 25.23 |
| 8 | 5.91 | 4.91 | 16.66 |
| 9 | 19.73 | 15.96 | 13.58 |
| 10 | 12.04 | 14.15 | 20.52 |
| 11 | 27.09 | 23.04 | 26.23 |
| 12 | 6.27 | 4.50 | 9.64 |
| 13 | 16.85 | 7.44 | 15.54 |
| 14 | 4.90 | 8.10 | 19.92 |
| 15 | 13.85 | 22.71 | 27.71 |
| 16 | 17.44 | 16.80 | 13.82 |
| **Mean** | 13.60 | **13.37** | 18.04 |

**Table Ic. One Video Leave Out (Stage 1 RMSE Values)**

| Video | XGBoost | Robust LMM | Baseline |
|---|---|---|---|
| 1 | 14.51 | 17.76 | 21.23 |
| 2 | 14.10 | 13.81 | 21.51 |
| 3 | 21.08 | 21.98 | 27.74 |
| **Mean** | **16.57** | 17.82 | 23.59 |

# APPENDIX J

## CROSS VALIDATION RESULTS FOR SATISFACTION PREDICTION

**Table Ja. Random Split (Stage 1 RMSE Values)**

| Split | XGBoost | Robust LMM | Baseline |
|---|---|---|---|
| 1 | 0.91 | 0.89 | 0.95 |
| 2 | 0.83 | 0.83 | 0.88 |
| 3 | 0.81 | 0.63 | 0.61 |
| 4 | 0.86 | 0.93 | 0.85 |
| 5 | 0.74 | 0.46 | 0.50 |
| **Mean** | 0.83 | **0.75** | 0.76 |

**Table Jb. One Person Leave Out (Stage 1 RMSE Values)**

| Person | XGBoost | Robust LMM | Baseline |
|---|---|---|---|
| 1 | 0.75 | 0.60 | 0.38 |
| 2 | 0.85 | 0.96 | 0.90 |
| 3 | 0.64 | 0.60 | 0.32 |
| 4 | 0.54 | 0.35 | 0.36 |
| 5 | 0.30 | 0.76 | 1.21 |
| 6 | 0.49 | 0.62 | 0.29 |
| 7 | 0.89 | 0.46 | 0.55 |
| 8 | 1.07 | 0.81 | 0.90 |
| 9 | 0.18 | 0.79 | 0.44 |
| 10 | 0.68 | 0.89 | 1.01 |
| 11 | 1.30 | 1.55 | 1.32 |
| 12 | 0.81 | 0.66 | 0.37 |
| 13 | 1.02 | 0.46 | 0.49 |
| 14 | 1.05 | 1.28 | 1.24 |
| 15 | 0.45 | 0.37 | 0.05 |
| 16 | 0.63 | 0.39 | 0.30 |
| **Mean** | 0.73 | 0.72 | **0.63** |

**Table Jc. One Video Leave Out (Stage 1 RMSE Values)**

| Video | XGBoost | Robust LMM | Baseline |
|---|---|---|---|
| 1 | 0.82 | 0.76 | 0.72 |
| 2 | 1.04 | 1.01 | 0.81 |
| 3 | 0.85 | 0.86 | 0.56 |
| **Mean** | 0.90 | 0.88 | **0.70** |

**Table Jd. Random Split (Stage 2 RMSE Values)**

| Split | Robust LMM | Baseline |
|---|---|---|
| 1 | 0.41 | 0.95 |
| 2 | 0.53 | 0.88 |
| 3 | 0.41 | 0.61 |
| 4 | 0.50 | 0.85 |
| 5 | 0.45 | 0.50 |
| **Mean** | **0.46** | 0.76 |

**Table Je. One Person Leave Out (Stage 2 RMSE Values)**

| Person | Robust LMM | Baseline |
|---|---|---|
| 1 | 0.45 | 0.38 |
| 2 | 0.21 | 0.90 |
| 3 | 0.31 | 0.32 |
| 4 | 0.20 | 0.36 |
| 5 | 0.29 | 1.21 |
| 6 | 0.48 | 0.29 |
| 7 | 0.48 | 0.55 |
| 8 | 0.30 | 0.90 |
| 9 | 0.75 | 0.44 |
| 10 | 0.55 | 1.01 |
| 11 | 0.34 | 1.32 |
| 12 | 0.56 | 0.37 |
| 13 | 0.58 | 0.49 |
| 14 | 0.55 | 1.24 |
| 15 | 0.37 | 0.05 |
| 16 | 0.39 | 0.30 |
| **Mean** | **0.43** | 0.63 |

**Table Jf. One Video Leave Out (Stage 2 RMSE Values)**

| Video | Robust LMM | Baseline |
|---|---|---|
| 1 | 0.40 | 0.72 |
| 2 | 0.42 | 0.81 |
| 3 | 0.46 | 0.56 |
| **Mean** | **0.43** | 0.70 |

# APPENDIX K

## CROSS VALIDATION RESULTS FOR ENGAGEMENT PREDICTION

**Table Ka. Random Split (Stage 1 RMSE Values)**

| Split | XGBoost | Robust LMM | Baseline |
|---|---|---|---|
| 1 | 1.18 | 1.50 | 1.30 |
| 2 | 1.55 | 1.42 | 1.44 |
| 3 | 0.99 | 0.89 | 1.11 |
| 4 | 1.40 | 1.17 | 1.19 |
| 5 | 1.24 | 0.91 | 0.97 |
| **Mean** | 1.27 | **1.18** | 1.20 |

**Table Kb. One Person Leave Out (Stage 1 RMSE Values)**

| Person | XGBoost | Robust LMM | Baseline |
|---|---|---|---|
| 1 | 1.52 | 1.11 | 0.72 |
| 2 | 1.39 | 1.42 | 0.88 |
| 3 | 1.61 | 1.99 | 0.84 |
| 4 | 1.15 | 0.71 | 0.79 |
| 5 | 0.56 | 0.20 | 0.39 |
| 6 | 1.21 | 1.08 | 0.80 |
| 7 | 0.64 | 0.65 | 1.04 |
| 8 | 1.20 | 1.01 | 0.76 |
| 9 | 0.47 | 0.25 | 0.77 |
| 10 | 1.26 | 1.21 | 1.22 |
| 11 | 2.18 | 2.25 | 2.28 |
| 12 | 1.13 | 0.91 | 0.26 |
| 13 | 2.16 | 1.61 | 1.30 |
| 14 | 1.78 | 1.65 | 1.89 |
| 15 | 1.53 | 1.33 | 0.92 |
| 16 | 0.79 | 0.71 | 0.52 |
| **Mean** | 1.29 | 1.13 | **0.96** |

**Table Kc. One Video Leave Out (Stage 1 RMSE Values)**

| Video | XGBoost | Robust LMM | Baseline |
|---|---|---|---|
| 1 | 1.76 | 1.69 | 1.54 |
| 2 | 1.18 | 1.21 | 1.17 |
| 3 | 1.30 | 1.34 | 1.11 |
| **Mean** | 1.41 | 1.41 | **1.27** |

**Table Kd. Random Split (Stage 2 RMSE Values)**

| Split | Robust LMM | Baseline |
|---|---|---|
| 1 | 1.15 | 1.30 |
| 2 | 1.04 | 1.44 |
| 3 | 0.93 | 1.11 |
| 4 | 1.20 | 1.19 |
| 5 | 0.75 | 0.97 |
| **Mean** | **1.02** | 1.20 |

**Table Ke. One Person Leave Out (Stage 2 RMSE Values)**

| Person | Robust LMM | Baseline |
|---|---|---|
| 1 | 0.93 | 0.72 |
| 2 | 0.58 | 0.88 |
| 3 | 1.44 | 0.84 |
| 4 | 0.20 | 0.79 |
| 5 | 1.37 | 0.39 |
| 6 | 1.15 | 0.80 |
| 7 | 0.51 | 1.04 |
| 8 | 0.81 | 0.76 |
| 9 | 0.20 | 0.77 |
| 10 | 1.07 | 1.22 |
| 11 | 1.36 | 2.28 |
| 12 | 1.82 | 0.26 |
| 13 | 1.33 | 1.30 |
| 14 | 1.30 | 1.89 |
| 15 | 1.27 | 0.92 |
| 16 | 0.74 | 0.52 |
| **Mean** | 1.01 | **0.96** |

**Table Kf. One Video Leave Out (Stage 2 RMSE Values)**

| Video | Robust LMM | Baseline |
|---|---|---|
| 1 | 1.46 | 1.54 |
| 2 | 0.85 | 1.17 |
| 3 | 1.19 | 1.11 |
| **Mean** | **1.17** | 1.27 |

## CURRICULUM VITAE

### PERSONAL INFORMATION

Surname, Name: Yürüm, Ozan Raşit

Nationality: Turkey (TC)

Date and Place of Birth: August 21, 1988, Bakırköy

Marital Status: Married

E-mail:

### EDUCATION

| Degree | Institution | Year of Graduation |
|---|---|---|
| Ph.D. | Information Systems, METU | 2021 |
| M.Sc. | Information Systems, METU | 2014 |
| B.Sc. | Computer Education and Instructional Technology, METU | 2011 |

### WORK EXPERIENCE

| Year | Place | Enrollment |
|---|---|---|
| 2021-present | Izmir Institute of Technology | Instructor |
| 2019-2021 | Izmir University of Economics | Instructor |
| 2011-2019 | Middle East Technical University | Research Assistant |

### FOREIGN LANGUAGES

English (advanced), German (beginner)

## HONORS / GRANTS

Distinguished Paper Award
*43rd Euromicro Conference on Software Engineering and Advanced Applications (SEAA 2017), Vienna*

Graduate Student Grant
*The Scientific and Technological Research Council of Turkey (TÜBİTAK)*

Graduation in 2nd Rank
*Computer Education and Instructional Technology, Middle East Technical University*

High Honor Student
*Computer Education and Instructional Technology, Middle East Technical University*

## PROJECTS

| Date | Title | Type | Enrollment |
|---|---|---|---|
| 01.10.2018-01.10.2019 | TeachUP | Erasmus+ | Tutor |
| 01.04.2014-26.01.2017 | Agile Maturity Model and Agility Assessment Tool | TÜBİTAK 1001 | Researcher |

## SEMINARS AND CLASSES

| Class | Date | Title |
|---|---|---|
| Certified Software Size Measurer | 03.2015 | COSMIC |
| Computer Networks and Service Infrastructure | 08.2012 | Intensive ERASMUS Programme (72 hours) |
| Software and Database Expertise (Certificate of Excellence) | 08.2010 | Bilge Adam Information Technologies Academy (300 hours) |

## PUBLICATIONS

Yürüm, O. R., Taskaya-Temizel, T., & Yıldırım, S. (2021). The Impact of Video Clickstream Behaviors on Academic Performance. *Journal of Computing in Higher Education* [Submitted].

Yürüm, O. R., Yıldırım, S., & Taskaya-Temizel, T. (2021). An Intervention Framework for Developing Interactive Video Lectures based on Video Clickstream Behavior: A Quasi-Experimental Evaluation. *Interactive Learning Environments* [Submitted].

Yürüm, O. R., & Yıldırım, S. (2020). Video Analitikler. In H. Yurdagül, T. Güyer, & S. Yıldırım (Eds.), *Eğitsel Veri Madenciliği ve Öğrenme Analitikleri* (pp. 185-202). Anı Yayıncılık.

Yürüm, O. R., Demirörs, O., & Rabhi, F. (2018). A Comprehensive Evaluation of Agile Maturity Self- Assessment Surveys. In I. Stamelos, R. O'Connor, T. Rout, & A. Dorling (Eds.), *Software Process Improvement and Capability Determination. SPICE 2018* (pp. 300-315). Springer. https://doi.org/10.1007/978-3-030-00623-5_21

Yürüm, O. R., & Kilis, S. (2018, May 2-4). *University Students' Self-efficacy for and Attitude toward Computer Programming and Their Relationship* [Conference presentation]. 12th International Computer and Instructional Technologies Symposium (ICITS).

Yürüm, O. R., & Demirörs, O. (2017). Agile Maturity Self-Assessment Surveys: A Case Study. In M. Felderer, H. Holmström Olsson, & A. Skavhaug (Eds.), *43rd Euromicro Conference on Software Engineering and Advanced Applications (SEAA)* [pp. 392-399]. IEEE. https://doi.org/10.1109/SEAA.2017.58

Yürüm, O. R., Top, Ö. Ö., & Demirörs, O. (2017). Assessing Software Processes over a New Generic Software Process Assessment Tool. *The Collegium of Economic Analysis Annals, 43*, 1-30 https://ideas.repec.org/a/sgh/annals/i43y2017plip.30.html

Yürüm, O. R., & Arslan, O. (2017, October 11-13). *A Review on Acceptance Models for Wearable Technologies* [Conference presentation]. 5th International Instructional Technologies Teacher Education Symposium (ITTES).

Yürüm, O. R., & Kilis, S. (2017, October 11-13). *Kırsal Kesimdeki Eğitim Fakültesi Öğrencilerinin Bilgisayara Karşı Tutumları* [Conference presentation]. 5. Uluslararası Öğretim Teknolojileri ve Öğretmen Eğitimi Sempozyumu (ITTES).

Yürüm, O. R. (2014). *GSPA: A Generic Software Process Assessment Tool* [Master's thesis, Middle East Technical University, Ankara, Turkey]. http://etd.lib.metu.edu.tr/upload/12617934/index.pdf

Yürüm, O. R., Top, Ö. Ö., Ertuğrul, A. M., & Demirörs, O. (2014). *Yazılım Süreç Değerlendirme Araçlarının Karşılaştırılması: Bir Çoklu Durum Çalışması*. In Y. Yesilada, A. Betin Can, & O. Demirörs (Eds.), *Proceedings of the 8th Turkish National Software Engineering Symposium* (pp. 360-371). CEUR-WS. http://ceur-ws.org/Vol-1221/35_Bildiri.pdf