

**UNIVERZA NA PRIMORSKEM
PEDAGOŠKA FAKULTETA**

**DOKTORSKA DISERTACIJA
NEŽKA SAJINČIČ**

KOPER 2024

UNIVERZA NA PRIMORSKEM

PEDAGOŠKA FAKULTETA

Doktorski študijski program tretje stopnje

Edukacijske vede

Doktorska disertacija

**AUDITORY EMOTIONAL DESIGN IN
MULTIMEDIA LEARNING: EDUCATIONAL
VIDEOS ON WOOD AS A BUILDING MATERIAL
SLUŠNO-ČUSTVENO OBLIKOVANJE PRI
VEČPREDSTAVNOSTNEM UČENJU:
IZOBRAŽEVALNI VIDEOPOSNETKI O LESU KOT
GRADBENEM MATERIALU**

Nežka Sajinčič

Koper 2024

Mentorica: prof. dr. Petra Dolenc

Somentorica:

izr. prof. dr. Anna Małgorzata Sandak

ACKNOWLEDGMENTS

First and foremost, I am deeply grateful to my advisor, Full Prof. Petra Dolenc, PhD, for taking me under your wing. Your unwavering support, both academically and emotionally, has been indispensable for me as a researcher and as a person. Thank you for believing in me, guiding me with optimism and determination, and always being in my corner.

I would also like to express my profound gratitude to my co-supervisor, Assist. Prof. Anna Sandak, PhD. You have been a reliable mentor throughout my journey, introducing me to the research community, including me in projects, providing numerous ideas, and being an inexhaustible source of enthusiasm.

I am also grateful to Full Prof. Andreja Istenič, PhD, for helping me find my path as a researcher.

Special thanks to all the research participants, and a heartfelt thank you to everyone who helped in recruiting them. This work would not have been possible without you. Listed in alphabetical order: Assist. Prof. Danijel Bratina, PhD; Assoc. Prof. Mateja Dovjak, PhD; Sara Fabjan; Assoc. Prof. Armand Faganel, PhD; Assist. Prof. Mojca Gregorski, PhD; Full Prof. Eric Hansen, PhD; Niki Hrovatin, PhD; Full Prof. Manja Kitek Kuzman, PhD; Assist. Prof. Borut Kodrič, PhD; Assist. Prof. Nina Krmac, PhD; Full Prof. Andreja Kutnar, PhD; Full Prof. Anders Qvale Nyrud, PhD; Assist. Prof. Rok Prislan, PhD; Assoc. Prof. Mariapaola Riggio, PhD; Matic Sašek; Sr. Instr. Suzana Sedmak, MSc; Full Prof. Kristina Sepčič, PhD; Blaž Simčič, PhD; Amy Simmons, MSc; Assist. Prof. Iztok Šušteršič, PhD; Črtomir Tavzes, PhD; Aleksandar Tošić, PhD.

I am thankful to my colleagues and the management at InnoRenew CoE for their support and collaboration.

To my friends and family, who stood by me through everything, offering a listening ear and constant support.

Finally, I would like to give my utmost thanks to Assist. Prof. Dean Lipovac, PhD. Your support throughout this adventure has been immeasurable. I am forever grateful for our insightful conversations, your constructive feedback, invaluable encouragement, and the countless ways you have helped me reach where I am today.

This research was supported by the European Commission by funding the InnoRenew project (Grant Agreement #739574) under the H2020 Widespread-2-Teaming programme and by the Republic of Slovenia (investment funding from the

Republic of Slovenia and the European Regional Development Fund). Part of this work was also conducted during the ClickDesign and WoodLCC projects, which were supported under the umbrella of ERA-NET Cofund ForestValue. Funding was also provided by the European Research Council (ERC) for the ARCHI-SKIN project (Grant #101044468).



IZJAVA O AVTORSTVU

Podpisani/a Nežka Sajinčič vpisna številka 98193011 študijski program Edukacijske vede

izjavljam, da je doktorska disertacija z naslovom Auditory emotional design in multimedia learning: Educational videos on wood as a building material / Slušno-čustveno oblikovanje pri večpredstavnostnem učenju: Izobraževalni videoposnetki o lesu kot gradbenem materialu pod mentorstvom prof. dr. Petre Dolenc in somentorstvom izr. prof. dr. Anne Małgorzate Sandak

- rezultat lastnega raziskovalnega dela,
- da so rezultati korektno navedeni,
- da nisem kršil/a avtorskih pravic in intelektualne lastnine drugih in
- da je elektronska različica, ki sem jo oddal/a, istovetna tiskani različici.

Zaradi naslednjih utemeljenih razlogov:

_____prosim

za odlog javne objave disertacije v repozitoriju Dissertations and Thesis (Proquest) in dLib.si (NUK) za dobo _____ (največ 3) let.

Podpis mentorja: _____

(samo v primeru, če delo ne sme biti javno dostopno)

Podpis odgovorne osebe naročnika in žig: _____

(samo v primeru, če delo ne sme biti javno dostopno)

Kraj in datum: _____ Podpis avtorja/ice: _____.

Izjavljam, da je mentor seznanjen z indeksom podobnosti doktorske disertacije, ki je 1,28%.

Kraj in datum: _____ Podpis avtorja/ice: _____

IZVLEČEK

Zavedanje javnosti o ukrepih proti podnebnim spremembam, kot je gradnja z lesom, je izredno pomembno pri doseganju ciljev trajnostnega razvoja. Priročno orodje za širjenje znanja na to temo so izobraževalni videoposnetki. Medtem ko so se raziskave o videoposnetkih do sedaj osredotočale predvsem na kognitivne in učne rezultate, so čustveni procesi pritegnili pozornost raziskovalcev šele pred kratkim. Dodajanje elementov z namenom sprožanja čustev, ki spodbujajo učenje, je zato zanimiv koncept, ki zahteva nadaljnje raziskovanje, zlasti v kontekstu slušnih elementov učnih gradiv. Poleg tega se raziskave o oblikovanju večpredstavnostnih vsebin osredotočajo predvsem na gradiva v maternem jeziku, kar ne odraža sodobnega načina učenja preko spleta, kjer je večina izobraževalnih vsebin dostopnih v tujem jeziku, predvsem angleščini. Disertacija preučuje učinke slušno-čustvenega oblikovanja in istojezičnih podnapisov na študente, ki se v tujem jeziku učijo o lesu kot gradbenem materialu. Izvedena sta bila dva eksperimenta, ki sta se osredotočala na intervenciji čustvenega oblikovanja – pripovedovalčev čustveni ton glasu in glasbo v ozadju – in njun vpliv na čustveno stanje, kognitivno procesiranje in učne dosežke študentov, zlasti tistih, katerih angleščina ni materni jezik. Študija 1 z 229 univerzitetnimi študenti je preučevala učinke čustvenega tona glasu pripovedovalca in podnapisov v jeziku posnetka. Rezultati so pokazali, da je čustveni ton pripovedovalca vplival na zaznavo študentov, vendar ni bistveno vplival na kognitivno obremenitev ali učne dosežke. Nadaljnja analiza je razkrila, da je navdušeni ton koristil študentom z nižjim razumevanjem angleščine, oviral pa tiste z višjim znanjem angleščine. Podnapisi so zmanjšali nivo zunanje kognitivne obremenitve in izboljšali odloženi transfer, zlasti pri udeležencih z nižjim znanjem angleščine. Študija 2, ki je vključevala 307 študentov, je preverjala vpliv glasbe v ozadju z različnimi stopnjami aktivacije. Ugotovitve so pokazale, da je umirjena glasba zmanjšala negativna aktivacijska čustva in izboljšala samooceno uspešnosti na testu znanja, vendar ni pomembno vplivala na nivo kognitivne obremenitve ali objektivne učne dosežke. Učinki glasbe v ozadju so se znova razlikovali glede na jezikovno znanje udeležencev, kar poudarja pomen individualnih razlik in specifičnega vpliva čustvenih oblikovalskih intervencij. Disertacija prispeva k oblikovanju smernic za ustvarjanje učinkovitih izobraževalnih videoposnetkov s poudarkom na slušno-čustvenem oblikovanju ob upoštevanju potreb učencev, ki se učijo v tujem jeziku.

Ključne besede: večpredstavnostno učenje, slušno-čustveno oblikovanje, glas predavatelja, čustveni ton, glasba v ozadju, istojezični podnapisi, učenje v tujem jeziku, les kot gradbeni material.

ABSTRACT

Auditory emotional design in multimedia learning: Educational videos on wood as a building material

Public awareness of climate change mitigation measures, such as wooden construction, is crucial for achieving sustainable development. Creating educational videos is an effective way to disseminate this knowledge. While research on educational videos has focused on cognitive and learning outcomes, the role of emotional processes has only recently gained attention. Incorporating features that induce emotions to enhance learning needs further exploration, particularly in auditory elements. Additionally, most multimedia learning research focuses on native language content, which doesn't reflect the trend of online learning in foreign languages, especially English. This dissertation examines the effects of auditory emotional design and same-language subtitles on students learning about wood as a building material in a foreign language. Two between-subjects experiments were conducted, focusing on emotional design interventions – specifically, the narrator's emotional tone and background music – and their impact on learners' affective, cognitive, and learning outcomes, especially for non-native English speakers. Study 1, with 229 university students, investigated the effects of the narrator's emotional tone conveyed through voice and same-language subtitles. Results showed that while the narrator's emotional tone influenced learners' perceptions, it did not significantly affect cognitive load or learning outcomes. However, splitting the sample by English proficiency revealed that the enthusiastic tone benefited learners with lower proficiency but hindered those with higher proficiency. Subtitles reduced extraneous cognitive load and improved delayed transfer, especially for learners with lower English proficiency. Study 2, involving 307 students, assessed the impact of background music with different activation levels. Findings indicated that calm music reduced negative emotions and improved self-evaluated test performance but did not significantly influence cognitive processing or objective learning outcomes. The effects of background music again varied with participants' language proficiency, highlighting the importance of individual differences and the nuanced impact of emotional design interventions. This dissertation contributes to guidelines for creating effective educational videos by emphasizing emotional design, auditory interventions, and the needs of learners using a non-native language.

Key words: multimedia learning, auditory emotional design, instructor's voice, emotional tone, background music, same-language subtitles, learning in a foreign language, wood as a building material.

CONTENTS

1 INTRODUCTION.....	1
1.1 Wood as a sustainable and healthy building material	1
1.2 Online educational videos as a tool for lifelong learning	3
2 THEORETICAL FRAMEWORK.....	6
2.1 Theories of learning	6
2.1.1 Cognitive Load Theory.....	6
2.1.2 Cognitive Theory of Multimedia Learning.....	7
2.1.3 The role of emotions in learning.....	9
2.1.4 Cognitive-Affective Theory of Learning with Media	11
2.1.5 Integrated Cognitive Affective Model of Learning with Multimedia	13
2.1.6 Cognitive Affective Model of E-Learning	13
2.2 Emotional design principles.....	14
2.2.1 Minimal manipulations and seductive details	15
2.2.2 Auditory emotional design in multimedia learning	17
2.2.2.1 Emotional tone in narrator's voice in educational videos	18
2.2.2.2 Music in educational videos	20
2.3 Learning in a foreign language with same-language subtitles.....	23
3 EMPIRICAL PART	26
3.1 Research problem, purpose, and objectives.....	26
3.2 Pre-study 1: Recognizing human emotion from the narrator's voice	28
3.2.1 Research hypotheses	28
3.2.2 Methodology	28
3.2.2.1 Research design	28
3.2.2.2 Participants	28
3.2.2.3 Materials	30
3.2.2.4 Instruments	30
3.2.2.5 Data collection	31
3.2.2.6 Data analysis	31
3.2.3 Results and interpretation.....	31

3.3 Study 1: Experiment on the effect of emotional tone in the narrator's voice and same-language subtitles	34
3.3.1 Research hypotheses	34
3.3.2 Methodology	34
3.3.2.1 Research design	34
3.3.2.2 Participants	35
3.3.2.3 Materials	37
3.3.2.4 Instruments	39
3.3.2.5 Data collection.....	48
3.3.2.6 Data analysis	49
3.3.3 Results and interpretation	51
3.3.3.1 Groups' description and comparison	51
3.3.3.2 Emotional tone of the narrator	57
3.3.3.3 Same-language subtitles.....	94
3.3.3.4 Interactions	122
3.3.3.5 Limitations and implications.....	130
3.4 Pre-studies 2: Music rating	132
3.4.1 Research hypotheses	132
3.4.2 Methodology	132
3.4.2.1 Research design	132
3.4.2.2 Participants	133
3.4.2.3 Material	134
3.4.2.4 Instruments	135
3.4.2.5 Data collection.....	136
3.4.2.6 Data analysis	136
3.4.3 Results and interpretation	136
3.5 Study 2: Experiment on the effect of background music	139
3.5.1 Research hypotheses	139
3.5.2 Methodology	139

3.5.2.1 Research design	139
3.5.2.2 Participants	139
3.5.2.3 Material	143
3.5.2.4 Instruments	144
3.5.2.5 Data collection	152
3.5.2.6 Data analysis	154
3.5.3 Results and interpretation	156
3.5.3.1 Groups' description and comparison	156
3.5.3.2 Background music and video perception	170
3.5.3.3 Emotional outcomes	175
3.5.3.4 Cognitive outcomes	194
3.5.3.5 Learning	198
3.5.3.6 Additional analyses	213
3.5.3.6 Limitations and implications	223
4 CONCLUSIONS	226
5 REFERENCES	231
SLUŠNO-ČUSTVENO OBLIKOVANJE PRI VEČPREDSTAVNOSTNEM UČENJU: IZOBRAŽEVALNI VIDEOPOSNETKI O LESU KOT GRADBENEM MATERIALU	248
KAZALO VSEBINE	248
POVZETEK VSEBINE IN UGOTOVITEV DOKTORSKE DISERTACIJE	252
6 APPENDIX	262
6.1 Appendix 1: Videos with their duration and corresponding URLs	262
6.2 Appendix 2: Pitch analysis of audio segments taken from the videos with enthusiastic and calm narrations	264
6.3 Appendix 3: Correlation matrix (Pearson <i>r</i>) of Study 1 outcome variables	268
6.4 Appendix 4: Songs used in Pre-study 2	270
6.5 Appendix 5: Pre-test questions with answers (correct in bold) in Slovene, English, and Norwegian	272
6.6 Appendix 6: Retention and transfer post-test questions with answers (correct in bold) in Slovene, English, and Norwegian	275
6.7 Appendix 7: Normality and homogeneity test results for Study 1 outcomes	286

6.8 Appendix 8: Descriptive statistics for main outcomes by low proficiency narrator group	290
6.9 Appendix 9: ANCOVA comparisons by low proficiency narrator group	294
6.10 Appendix 10: Descriptive statistics for main outcomes by high proficiency narrator group	297
6.11 Appendix 11: ANCOVA comparisons by high proficiency narrator group ...	301
6.12 Appendix 12: Descriptive statistics for main outcomes by low proficiency SLS group	304
6.13 Appendix 13: ANCOVA comparisons by low proficiency SLS group	308
6.14 Appendix 14: Descriptive statistics for main outcomes by high proficiency SLS group	311
6.15 Appendix 15: ANCOVA comparisons by high proficiency SLS group	315
6.16 Appendix 16: Two-way ANCOVA comparisons for low proficiency group – Study 1	318
6.17 Appendix 17: Two-way ANCOVA comparisons for high proficiency group – Study 1	322
6.18 Appendix 18: List of study programs for participants in Study 2	326
6.19 Appendix 19: Correlation matrix (Pearson r) of Study 2 outcome variables	327
6.20 Appendix 20: Normality and homogeneity test results for Study 2 outcomes	329
6.21 Appendix 21: Descriptive statistics by proficiency – Study 2	332
6.22 Appendix 22: ANCOVA comparisons for low proficiency group – Study 2 .	341
6.23 Appendix 23: ANCOVA comparisons for high proficiency group – Study 2	344
6.24 Appendix 24: Descriptive statistics by wood science familiarity – Study 2 .	347
6.25 Appendix 25: ANCOVA comparisons for wood science group – Study 2 ...	356
6.26 Appendix 26: ANCOVA comparisons for non-wood science group – Study 2	359

LIST OF TABLES

Table 1: Country of participants ($n = 209$).....	29
Table 2: Educational level and status of participants ($n = 209$).....	29
Table 3: Paired t -tests comparing enthusiastic and calm ratings in enthusiastic and calm videos with other emotions	32
Table 4: Demographics divided by country and in total.....	36
Table 5: Item difficulty indexes of pre-test questions in Study 1.....	41
Table 6: Item difficulty indexes and confidence levels of correct responses on post-test questions in Study 1	42
Table 7: Learners' characteristics and descriptive statistics for variables before watching the videos divided by enthusiastic and calm conditions	53
Table 8: Learners' characteristics and descriptive statistics for variables before watching the videos divided by group without and with SLS	54
Table 9: t -tests, normality, and homogeneity tests' results comparing the enthusiastic vs. calm narrator groups and the no SLS vs. SLS groups	55
Table 10: Descriptive statistics for variables related to recognizing the narrator's emotion for enthusiastic and calm narrator groups.....	57
Table 11: Comparisons of the enthusiastic and calm narrator groups on discrete emotions, pleasantness, and activation level rating of the narrator using t -tests ...	58
Table 12: ANCOVA and post-hoc comparisons of the enthusiastic and calm voice groups on discrete emotions, pleasantness and activation level rating of the narrator.....	59
Table 13: Descriptive statistics for API components for enthusiastic and calm narrator groups	61
Table 14: Comparison of the enthusiastic and calm voice groups on the Facilitating learning, Human-likeness, Credibility, and Engaging variables using Welch's t -tests	62
Table 15: ANCOVA comparisons with five covariates of the enthusiastic and calm voice groups on the Facilitating learning, Human-likeness, Credibility, and Engaging variables.....	63
Table 16: Descriptive statistics of PANAVA-KS values and change score for enthusiastic and calm narrator groups.....	65
Table 17: Baseline and post-intervention differences in the PANAVA-KS measures for the enthusiastic narrator and the calm narrator groups separately.....	66
Table 18: Descriptive statistics of activation level and valence measurements and change score for enthusiastic and calm narrator groups.....	68

Table 19: Baseline and post-intervention differences in the activation level and valence measures for the enthusiastic narrator and the calm narrator groups	69
Table 20: ANCOVA comparisons of the enthusiastic and calm voice groups on activation level and valence items	71
Table 21: ANCOVA comparisons with five covariates of the enthusiastic and calm voice groups on activation level and valence items	72
Table 22: Descriptive statistics for the learners' experience questions for the enthusiastic and calm voice groups	74
Table 23: Comparison of the enthusiastic and calm voice groups on several variables on the learners' experience using Welch's <i>t</i> -tests	74
Table 24: ANCOVA comparisons with six covariates of the enthusiastic and calm voice groups on several variables on the learners' experience	75
Table 25: Descriptive statistics of the cognitive load questionnaire for enthusiastic and calm narrator groups	76
Table 26: Descriptive statistics of the mental effort ratings for enthusiastic and calm narrator groups	77
Table 27: Comparison of the enthusiastic and calm voice groups on mental effort.....	78
Table 28: ANCOVA comparisons with six covariates of the enthusiastic and calm voice groups on mental effort.....	79
Table 29: Descriptive statistics of the learning outcomes from the immediate part of the experiment for enthusiastic and calm narrator groups	80
Table 30: Comparison between the enthusiastic and calm voice groups on various learning variables in the immediate part of the experiment using Welch's <i>t</i> -tests .	81
Table 31: ANCOVA comparisons with six covariates of the enthusiastic and calm voice groups on various learning variables in the immediate part of the experiment	83
Table 32: Quade test comparisons with six covariates of the enthusiastic and calm voice groups on various certainty level variables in the immediate part of the experiment	84
Table 33: Descriptive statistics of the learning outcomes from the delayed part of the experiment for enthusiastic and calm narrator groups (<i>N</i> = 94).....	85
Table 34: Comparison between the enthusiastic and calm voice groups on various learning variables in the delayed part of the experiment using Student's <i>t</i> -tests ..	86
Table 35: ANCOVA comparisons with six covariates of the enthusiastic and calm voice groups on various learning variables in the delayed part of the experiment	87
Table 36: Descriptive statistics of the learning outcomes from the immediate (<i>N</i> = 224) and delayed part of the experiment (<i>N</i> = 94).....	88

Table 37: Pairwise comparisons of the learning variables in the immediate and delayed parts of the experiment with normality test.....	89
Table 38: Descriptive statistics for variables related to recognizing the narrator's emotion for groups with and without SLS	95
Table 39: Comparison of the groups with and without SLS on discrete emotions, pleasantness and activation level rating of the narrator using <i>t</i> -tests	95
Table 40: ANCOVA comparisons with six covariates of the groups without and with SLS on discrete emotions, pleasantness and activation level rating of the narrator.....	96
Table 41: Descriptive statistics for API components for group without and with SLS ...	97
Table 42: Comparison of the groups without and with SLS on the Facilitating learning, Human-likeness, Credibility, and Engaging variables using <i>t</i> -tests	97
Table 43: ANCOVA comparisons with five covariates of the groups without and with SLS on the Facilitating learning, Human-likeness, Credibility, and Engaging variables	98
Table 44: Descriptive statistics of PANAVA-KS values and change score for groups without and with SLS	99
Table 45: Baseline and post-intervention differences in the PANAVA-KS measures for the group without SLS and the group with SLS separately	100
Table 46: Descriptive statistics of activation level and valence measurements and change score for groups without and with SLS	102
Table 47: Baseline and post-intervention differences in the activation level and valence measures for groups without and with SLS.....	103
Table 48: ANCOVA comparisons of the groups with and without SLS on activation level and valence items	104
Table 49: ANCOVA comparisons with five covariates of the groups with and without SLS on activation level and valence items	105
Table 50: Descriptive statistics for the learners' experience questions for groups without and with SLS	107
Table 51: Comparison of the groups without and with SLS on several variables on the learners' experience	108
Table 52: ANCOVA comparisons with six covariates of the groups without and with SLS on several variables on the learners' experience	109
Table 53: Descriptive statistics of the cognitive load questionnaire for groups without and with SLS	110
Table 54: Comparison of the groups without and with SLS on cognitive load	110
Table 55: ANCOVA comparisons with six covariates of the groups without and with SLS on cognitive load	111

Table 56: Descriptive statistics of the mental effort ratings for groups without and with SLS	112
Table 57: Comparison of the groups without and with SLS on mental effort	112
Table 58: ANCOVA comparisons with six covariates of the groups without and with SLS on mental effort	113
Table 59: Descriptive statistics of the learning outcomes from the immediate part of the experiment for groups without and with SLS.....	114
Table 60: Comparison between the groups without and with SLS on various learning variables in the immediate part of the experiment using Welch's <i>t</i> -tests.....	115
Table 61: ANCOVA comparisons with six covariates of the groups with and without SLS on various learning variables in the immediate part of the experiment.....	116
Table 62: Descriptive statistics of the learning outcomes from the delayed part of the experiment for groups without and with SLS.....	117
Table 63: Comparison between the groups without and with SLS on various learning variables in the delayed part of the experiment using Welch's <i>t</i> -tests	118
Table 64: ANCOVA comparisons with six covariates of the groups with and without SLS on various learning variables in the delayed part of the experiment.....	120
Table 65: Two-way ANCOVA comparisons of the instructor perception variables, together with homogeneity tests	122
Table 66: Two-way ANCOVA comparisons of the emotional outcomes variables, together with homogeneity tests	124
Table 67: Two-way ANCOVA comparisons of the cognitive outcomes variables, together with homogeneity tests	125
Table 68: Two-way ANCOVA comparisons of the main learning variables, together with homogeneity tests	126
Table 69: Demographics from both pre-studies	133
Table 70: Songs with their corresponding energy level and emotional tone ratings ..	137
Table 71: Paired <i>t</i> -tests comparing valence and activation level ratings of Song F and O with baseline ratings and change scores of Song F and O between themselves	138
Table 72: Demographics divided by university and in total	140
Table 73: Item difficulty indexes of pre-test questions in Study 2.....	145
Table 74: Item difficulty indexes and confidence levels of correct responses on post-test questions.....	146
Table 75: Learners' characteristics and descriptive statistics for variables before watching the videos divided by educational field.....	158
Table 76: Comparisons of learners' characteristics before watching the videos divided by educational field using Welch's <i>t</i> -tests.....	160

Table 77: Learners' characteristics and descriptive statistics for variables before watching the videos divided by group	162
Table 78: One-way ANOVA comparisons of the learners' characteristics and variables before watching the videos between experimental groups.....	166
Table 79: Games-Howell post-hoc tests of three learners' characteristics and variables before watching the videos between experimental groups.....	167
Table 80: Descriptive statistics for the two video perception variables divided by group	173
Table 81: Post-hoc comparisons for Video pleasantness	174
Table 82: ANCOVA post-hoc comparisons for Video pleasantness.....	175
Table 83: Descriptive statistics of PANAVA-KS values and change score divided by group	176
Table 84: Baseline and post-intervention differences in the PANAVA-KS measures for the three experimental groups separately	177
Table 85: Post-hoc comparisons for Negative activation	179
Table 86: ANCOVA post-hoc comparisons for Negative activation.....	180
Table 87: Descriptive statistics of activation level and valence measurements and change score divided by group.....	181
Table 88: Baseline and post-intervention differences in the activation level and valence measures for the three experimental groups separately	183
Table 89: ANCOVA comparisons on activation level and valence items.....	185
Table 90: Post-hoc comparisons for variables Valence 1, 2, and 3.....	186
Table 91: ANCOVA/Quade comparisons (with multiple covariates) on activation level and valence items	187
Table 92: ANCOVA post-hoc comparisons for Valence 1, 2, 3, and average	188
Table 93: Descriptive statistics of the two interest variables divided by group	189
Table 94: Post-hoc comparisons for Situational interest	189
Table 95: ANCOVA post-hoc comparisons for Situational interest.....	190
Table 96: Descriptive statistics of the learners' experience variables divided by group	191
Table 97: Comparisons of the three groups on learners' experience variables	192
Table 98: Post-hoc comparisons for the Paying attention and More lessons like this variables.....	192
Table 99: ANCOVA comparisons with eleven covariates on the learners' experience variables.....	193
Table 100: ANCOVA post-hoc comparisons for the Paying attention, Enjoyment, and More lessons like this variables	194

Table 101: Descriptive statistics of the cognitive load questionnaire divided by group	195
Table 102: Comparisons of the three groups on learners' experience variables	195
Table 103: Descriptive statistics of the mental effort ratings divided by group	196
Table 104: ANOVA comparisons of the three groups on mental effort.....	197
Table 105: ANCOVA comparisons with eleven covariates on mental effort	198
Table 106: Descriptive statistics of the learning outcomes from the immediate part of the experiment divided by group.....	199
Table 107: ANOVA comparisons of the three groups on learning outcomes in the immediate part of the experiment	201
Table 108: Post-hoc comparisons for learning outcomes in the immediate part of the experiment.....	202
Table 109: ANCOVA comparisons with eleven covariates on learning variables in the immediate part of the experiment	203
Table 110: ANCOVA post-hoc comparisons for some of the learning outcome variables in the immediate part of the experiment.....	204
Table 111: Descriptive statistics of the learning outcomes from the delayed part of the experiment divided by group.....	206
Table 112: ANOVA comparisons of the three groups on learning outcomes in the delayed part of the experiment.....	208
Table 113: Post-hoc comparisons for Self-evaluated test performance	209
Table 114: ANCOVA comparisons with eleven covariates on learning variables in the delayed part of the experiment	209
Table 115: ANCOVA post-hoc comparisons for some of the learning outcome variables in the delayed part of the experiment.....	210
Table 116: Descriptive statistics of the learning outcomes from the immediate ($N = 304$) and delayed part of the experiment ($N = 118$).....	212
Table 117: Pairwise comparisons of the learning variables in the immediate and delayed parts of the experiment.....	213
Table 118: ANCOVA post-hoc comparisons for the lower proficiency group – Study 2	214
Table 119: ANCOVA post-hoc comparisons for the higher proficiency group – Study 2	216
Table 120: ANCOVA post-hoc comparisons for the wood science related educational programs	219
Table 121: ANCOVA post-hoc comparisons for the educational programs not related to wood science related.....	220

Table 122: Big five personality characteristics as covariates on main outcome variables	222
Table 123: Shapiro-Wilk's normality and Levene's homogeneity tests for Study 1 outcome variables before comparisons	286
Table 124: Descriptive statistics for the main outcome variables for enthusiastic and calm narrator groups on the lower English proficiency group (LexTALE < 63)	290
Table 125: ANCOVA comparisons between the enthusiastic and calm narrator on all main dependable variables on the lower English proficiency group (LexTALE < 63)	294
Table 126: Descriptive statistics for the main outcome variables for enthusiastic and calm narrator groups on the higher English proficiency group (LexTALE > 63)	297
Table 127: ANCOVA comparisons between the enthusiastic and calm narrator on all main dependable variables on the higher English proficiency group (LexTALE > 63)	301
Table 128: Descriptive statistics for the main outcome variables for the groups without and with SLS on the lower English proficiency group (LexTALE < 63)	304
Table 129: ANCOVA comparisons between the groups without and with SLS on all main dependable variables on the lower English proficiency group (LexTALE < 63) ...	308
Table 130: Descriptive statistics for the main outcome variables for the groups without and with SLS on the higher English proficiency group (LexTALE > 63)	311
Table 131: ANCOVA comparisons between the groups without and with SLS on all main dependable variables on the higher English proficiency group (LexTALE > 63) ..	315
Table 132: Two-way ANCOVA comparisons on all main dependable variables on the lower English proficiency group (LexTALE < 63), together with homogeneity tests	318
Table 133: Two-way ANCOVA comparisons on all main dependable variables on the higher English proficiency group (LexTALE > 63), together with homogeneity tests	322
Table 134: Shapiro-Wilk's normality and Levene's homogeneity tests for Study 2 outcome variables before ANOVAs	329
Table 135: Descriptive statistics for the main outcome variables divided by lower (LexTALE < 69) and higher (LexTALE > 69) English proficiency group – Study 2	332
Table 136: ANCOVA comparisons on all main dependable variables on the lower English proficiency group (LexTALE < 69) – Study 2	341
Table 137: ANCOVA comparisons on all main dependable variables in the higher English proficiency group (LexTALE > 69) – Study 2	344

Table 138: Descriptive statistics for the main outcome variables divided by lower study program familiarity with wood science – Study 2.....	347
Table 139: ANCOVA comparisons on all main dependable variables in participants from educational fields related to wood science – Study 2.....	356
Table 140: ANCOVA comparisons on all main dependable variables in participants from educational fields not related to wood science – Study 2.....	359

LIST OF FIGURES

Figure 1: The cognitive theory of multimedia learning (CTML) (Mayer, 2014).....	8
Figure 2: Model of core affect (Russell, 1980).	10
Figure 3: Watson and Tellegen's (1985) two-dimensional model (with Russell's model with grey dotted lines for comparison).	11
Figure 4: CATLM (Moreno, 2006).....	12
Figure 5: The Cognitive-Affective Model of E-Learning (Lawson et al., 2021b).....	14
Figure 6: The Yerkers-Dodson law (Teigen, 1994)	21
Figure 7: Averaged ratings of the enthusiastic videos.....	32
Figure 8: Averaged ratings of the calm videos.....	33
Figure 9: Experimental procedure of Study 1.....	48
Figure 10: Experimental procedure of Study 2.....	153
Figure 11: Frequency of studying with music in the background.....	168
Figure 12: Frequency of music genres listened to during studying.	169
Figure 13: Background music influence perception answers frequency divided by group.	171
Figure 14: Background music influence perception categories frequency divided by group.....	172

1 INTRODUCTION

1.1 Wood as a sustainable and healthy building material

The construction industry is a major economic sector in Europe, significantly influencing employment and economic growth. It directly employs about 18 million people and contributes to 9% of the European Union's GDP (European Commission, n.d. - a). Wood-based industries, a vital subset of this sector, support over 1 million jobs across approximately 184,000 enterprises throughout the EU (Jonsson et al., 2021). The escalating demands of Europe's growing population make the construction industry even more vital, as sufficient infrastructure needs to be developed to meet housing and service needs.

However, the construction industry poses large environmental challenges. Buildings are responsible for about 40% of energy consumption and nearly half of all CO₂ emissions within the EU (Bonoli et al., 2021; Clarke and Sahin-Dikmen, 2020), making a shift towards sustainable building approaches a pressing issue. Sustainable development is defined as "meeting the needs of the present without compromising the ability of future generations to meet their own needs" (United Nations, 2015). Sustainable practices within the construction sector can address both environmental impacts and demographic challenges, making it a key area for ongoing research, development, and innovation to improve material efficiency, reduce energy use, and ensure ecological and social well-being.

Within the material-intensive construction industry, wood holds a major promise in upholding sustainability goals. The unique characteristics of wood, including its high strength-to-weight ratio, thermal and acoustical insulation properties, and aesthetic versatility, make it a preferred choice for a wide range of construction applications (Asdrubali et al., 2017; Falk, 2009; Song et al., 2018).

Building with wood has significant environmental benefits, as wood is a natural and renewable resource that, when sourced responsibly, has a lower carbon footprint compared to non-renewable building materials (Tellnes et al., 2017). The sustainability of wood in construction is multifaceted, stemming from its role as a carbon sink, its energy-efficient production processes, and its extensive lifecycle. Wood sequesters carbon during its growth, and this carbon remains stored throughout the wood's use, contributing significantly to environmental conservation and climate change mitigation (Falk, 2009). As wood is both lightweight and strong, it is easier to work with and less

energy-intensive to transport than many other materials. Advancements in forestry management and wood processing technologies have expanded its uses beyond traditional framing and structural components, making it more convenient for modern building designs and eco-friendly construction practices (Corduban et al., 2012). Compared to conventional construction materials like concrete and steel, wood and engineered wood products have the ability to reduce carbon footprints, enhance energy efficiency, and minimize waste (Sandak et al., 2020).

As a natural material, wood has also been gaining attention for its potential positive impacts on human well-being and comfort in indoor environments. Research has increasingly focused on how wood and wooden materials influence the psychological and physiological states of people. The presence of wood in interiors has been linked to enhanced well-being of occupants (Alapieti et al., 2020). For example, some authors noted that wooden indoor environments induced more pleasant emotions, reduced feelings of fatigue, and supported physiological regulation compared to non-wooden settings (Zhang et al., 2016, 2017). Similar findings have been reported by Burnard and Kutnar (2015). The authors have reviewed existing studies on the psychophysiological effects of wood use in interiors, finding that environments with a higher presence of wood can lead to reduced stress responses, suggesting that wood might be an effective addition to indoor spaces to improve the well-being of building occupants (Burnard and Kutnar, 2015).

Despite the numerous advantages, building with wood does pose certain challenges, due to reasons such as susceptibility to biological degradation, fire risks, and environmental concerns related to deforestation. However, with sustainable forestry practices, enhanced treatment methods, and innovative construction techniques, these challenges can be mitigated.

The ongoing development of wood science and technology and advancements in treatment methods and construction techniques continue to expand the possibilities of wood as a sustainable, healthy, and versatile building material with enhanced durability and safety (Gan et al., 2019; Goldhahn et al., 2021; Hill and Dibdiakova, 2016; Jiang et al., 2018). As the construction industry progresses towards greener and more sustainable practices, the importance of wood is expected to grow, reinforcing its position as a key material in the future of construction.

However, even though the advantages of wood compared to other common building materials are substantial, these benefits are not universally recognized, either by the professional community or the general public. Research shows that consumers know

sustainable alternatives only superficially, but in-depth knowledge is associated with greater use of more sustainable options (Sajinčič et al., 2021). Due to this, it is becoming crucial to devise effective educational practices that can convey the myriads of important lessons about building with wood to professionals and laypeople alike, as comprehensive educational strategies are essential to catalyse the broader adoption of wood in modern construction practices.

Educational initiatives targeting professionals in the construction industry can demystify misconceptions about wood while promoting its advantages (Forest and Wood Products Australia, 2018). It is especially important to highlight newer and innovative techniques such as wood modification and engineered wood, such as cross-laminated timber and glue-laminated timber, which enhance wood's structural integrity and expand its applicability in high-rise buildings and large-span structures. These technologies not only facilitate the broader use of wood in diverse construction scenarios but also exemplify the advancements in wood engineering that align with modern architectural needs and sustainability goals (WoodSolutions, 2019).

The adoption of wood as a key construction material also requires public and policymaker education. Awareness campaigns and informational dissemination about the benefits of wood can lead to more supportive policies, including incentives for using sustainable materials and stricter regulations on materials with higher carbon footprints. This shift in policy can create a more favourable market for wood and encourage its use on a larger scale (Green Building Council, 2020).

1.2 Online educational videos as a tool for lifelong learning

Lifelong learning represents a shift in educational philosophy, suggesting that learning extends far beyond formal schooling into all aspects of life. In the modern knowledge economy, where new information is constantly emerging and job roles are continually evolving, the ability to learn throughout one's lifetime is critical for personal and professional development. Lifelong learning is not only about career advancement; it also enriches personal lives, fostering a culture of curiosity, self-improvement, and adaptability (Aspin and Chapman, 2000; Kind and Evans, 2015).

Online learning has emerged as a key enabler of lifelong learning. The role of online learning challenges traditional educational models by offering an alternative that emphasizes self-education, flexibility, and the sharing of knowledge through the Internet. This paradigm shift is especially important among Generation Z, who predominantly

derive knowledge from online sources and show a strong preference for mobile applications and video content over traditional learning forms (Szymkowiak et al., 2021).

Amidst the vast array of online learning tools, educational videos stand out as a particularly impactful resource. They have become an increasingly common and integral part of teaching and learning across various levels of education. Educational videos can break down barriers to learning, making content more engaging and less intimidating for individuals embarking on new areas of study. This is particularly relevant in fields that are either challenging or unfamiliar to the learner, as videos can present information in a more digestible and relatable format (Peters and Romero, 2019; Steffens, 2015).

With the advent of digital technology and platforms like YouTube, the accessibility and use of educational videos have soared, engaging millions of learners worldwide. However, despite their widespread usage, there remains a significant gap in the provision of clear, actionable guidelines for educators on how to create high-yield educational videos that meet the learning needs of learners (Krumm et al., 2021). There is a need for a structured framework for the development of educational videos that aligns with learning objectives and employs design guidelines focused on enhancing learning outcomes (Moussiades et al., 2019). This lack of clarity and understanding underscores the need for empirical evidence to inform the creation and use of educational videos, ensuring they are as effective and beneficial as possible.

While motivation to learn can be intrinsic, stemming from an individual's internal desire to learn and understand, or extrinsic, driven by external rewards and pressures (Ryan and Deci, 2000), educational videos have the potential to cater to both types, making learning more appealing. In educational settings, students may not always find the subject matter interesting or may lack prior knowledge. When intrinsic motivation is less pronounced, they may struggle to overcome challenges solely through desire and willpower. Educational videos, in this case, can make the content more accessible and less intimidating. For instance, videos that break down complex concepts into manageable parts, use clear and concise language, and include visual aids can help reduce cognitive load and make learning more manageable, thereby improving students' confidence and willingness to engage with the material (Choi and Johnson, 2005).

Creating educational videos that effectively motivate learners requires careful consideration of content, format, and presentation. In a 2015 study of YouTube's instructional videos, the authors found that popular videos were of higher production quality, had more static images, a combination of static and dynamic images, short on-screen texts, and background music, offered subtitles in various languages, included

less background noise, and had a faster speaking rate (ten Hove and van der Meij, 2015). However, popularity does not necessarily mean that these educational videos are effective. Videos should be designed to cater to both intrinsic and extrinsic motivators by making the content relevant, relatable, and applicable to real-world scenarios, thereby fostering a deeper understanding and appreciation of the subject matter (Brame, 2016).

Interest in the topic and motivation are individual differences that can significantly affect multimedia learning outcomes (Endres et al., 2020). Learners who are highly interested in a topic or who exhibit high levels of motivation are more likely to engage deeply with multimedia materials and experience more meaningful learning outcomes. Conversely, learners with low interest or motivation may require additional instructional support to achieve similar outcomes. In addition, like the characteristics of the learners, the properties of the learning content are also important. A recent meta-analysis revealed that the impact of multimedia learning interventions can vary depending on the subject area or instructional domain (Beege et al., 2023). Multimedia learning studies often purposefully introduce learning content that is not familiar to participants (e.g., Lehmann and Seufert, 2018; Liew and Tan, 2016). Prior knowledge is an important predictor of video lecture design, as people with less prior knowledge process information differently than more knowledgeable learners, and the optimal ways to present learning material differ between the two groups (Kalyuga et al., 2003). By choosing the study topic and participants where some participants will have higher prior knowledge and interest compared to others, studies can examine how these characteristics interact with the educational materials. Teaching about wood as a building material thus presents a unique set of challenges, as the topic may not inherently capture the interest of all learners, making it an ideal candidate for researching how learners perceive and interact with the subject and how to design learning videos that are more effective.

Additionally, as part of the European Year of Skills, the New European Bauhaus (European Commission, n.d. - b) introduces the NEB Academy, focusing on sustainable construction skills. The NEB Academy aims to accelerate up-skilling and re-skilling within the construction industry, facilitating the shift from a mineral-based, fossil fuel-dependent construction economy to a regenerative bio-economy and circular material reuse system. Within this context, studying the topic of wood as a building material, aligns with the broader goals of promoting sustainable practices and environmental awareness. By contributing to the NEB's initiatives, this research can enhance the effectiveness of education in sustainable construction and support the implementation of Green Deal principles by fostering a more informed and skilled workforce.

2 THEORETICAL FRAMEWORK

2.1 Theories of learning

2.1.1 Cognitive Load Theory

Cognitive Load Theory (CLT) provides important insights into how we can optimize educational experiences like learning videos to align with the human cognitive architecture. Originating in the 1980s, CLT has significantly influenced educational research and instructional design by highlighting the limited capacity of working memory, which plays a crucial role in the process of learning and problem-solving (Sweller, 1994). This concept is linked to Baddeley's model of working memory (Baddeley and Hitch, 1974), describing that working memory consists of multiple components – the central executive (functioning as the control center), the phonological loop and visuospatial sketchpad (handling verbal and visual-spatial information, respectively), and the episodic buffer (integrating information). While Baddeley's model explains the structure and function of working memory, CLT focuses on the effect working memory has on our cognitive processes, including learning and problem-solving.

In general, cognitive load refers to the total amount of mental effort or capacities used by the working memory. CLT delineates three types of cognitive load: intrinsic, extraneous, and germane.

Intrinsic cognitive load refers to the complexity inherent in the educational material itself. It is an unavoidable aspect of learning content that varies depending on the nature of the material being studied. For example, solving a complex algebraic equation inherently demands a higher cognitive load than memorizing a simple mathematical fact (Sweller, 1994; Sweller et al., 2011).

Extraneous cognitive load, on the other hand, relates to the way information is presented to learners. This type of cognitive load can be manipulated through instructional design. Poorly designed learning materials, such as those with unclear instructions or irrelevant information, increase the extraneous load and use up cognitive resources that could be better spent on processing relevant information (Mayer and Moreno, 2003).

Finally, germane cognitive load is associated with the cognitive effort directed towards understanding and creating new knowledge structures (schemas) and meaningful learning. This load is beneficial and represents the cognitive effort required

to process, construct, and automate schemas. Germane load is influenced by the instructional design that encourages learners to engage in meaningful learning activities that promote schema acquisition and automation (Sweller et al., 1998).

Each type of load plays a critical role in cognitive architecture and has implications for instructional design. Effective instructional design, according to CLT, should aim to avoid cognitive overload, manage intrinsic load, and optimize germane load.

CLT has provided an essential framework for understanding and improving learning and instruction, but significant work remains in applying its principles to the evolving landscape of educational technology and multimedia learning. As educational content increasingly shifts online and becomes more multimedia-rich, understanding how different media formats contribute to cognitive load becomes increasingly critical. The role of multimedia in managing or exacerbating cognitive load is a growing area of research within the framework of CLT, suggesting a need for ongoing investigation and adaptation of theory to new educational technologies (Mayer and Moreno, 2003).

2.1.2 Cognitive Theory of Multimedia Learning

Another extremely important learning theory that greatly influenced the literature on instructional design of learning materials, including educational videos, is the Cognitive Theory of Multimedia Learning (CTML). CTML revolves around the concept of multimedia instructional messages – presentations that consist of words (written or spoken) and visuals with the goal of promoting learning (Mayer, 2014). CTML is a cognitive processing theory that derives a lot of its' principles from CLT. Similarly to CLT, CTML posits that an effective multimedia educational design can enhance learning by aligning with the workings of the human cognitive system (Mayer and Moreno, 2003). CTML focuses not on technology but puts an emphasis on students themselves and the way their minds work, answering the question of how to adapt different learning environments to support the needs of human cognition (Mayer, 2014).

According to Mayer (2014), this theory builds on three main assumptions about human cognitive functioning, derived from empirical research on learning and cognition: dual channels, limited capacity, and active processing.

The dual channels assumption posits that humans process visual and auditory information through separate channels in the brain, allowing for more efficient handling of sensory information. This is supported by research suggesting that people have

distinct pathways for processing auditory and visual information, which can be optimized for more effective learning (Mayer, 2014; Mayer and Moreno, 2003).

The limited capacity assumption reflects the understanding that each cognitive channel can only process a certain amount of information at a time. This concept is essential for designing educational materials that do not overwhelm the learner's cognitive capabilities, thereby avoiding cognitive overload and enhancing learning efficiency (Mayer and Moreno, 2003).

The third assumption claims that active processing is crucial for meaningful learning. It occurs when learners actively engage in processing information, such as by organizing incoming information and integrating it with existing knowledge. Effective instructional design should facilitate these cognitive processes to enhance learning outcomes (Mayer, 2014).

Taken together, the CTML specifies five cognitive processes involved in multimedia learning (represented as arrows in the model in Figure 1):

- Selecting relevant words: learners choose relevant words from the presented text or narration.
- Selecting relevant images: learners select relevant images from the presented illustrations or graphics.
- Organizing verbal representation: selected words are organized into a coherent verbal representation.
- Organizing pictorial representation: selected images are organized into a coherent pictorial representation.
- Integrating pictorial and verbal representations: learners integrate the pictorial and verbal representations with their prior knowledge.

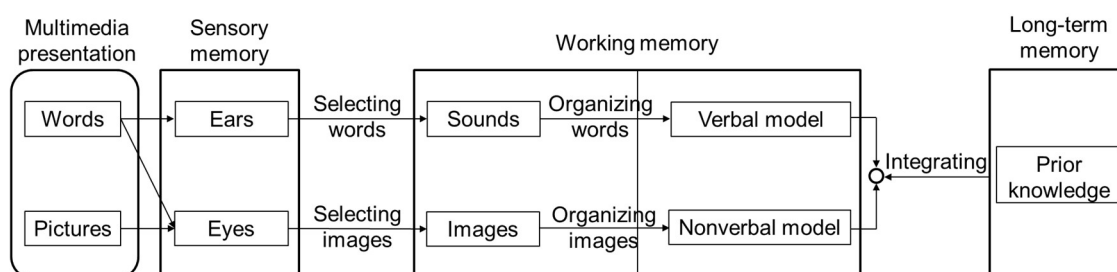


Figure 1: The cognitive theory of multimedia learning (CTML) (Mayer, 2014).

Based on these basic assumptions, research based on CTML has proposed several principles and strategies to help designing more effective multimedia learning materials, such as principles aimed at 1) reducing extraneous processing, allowing learners to

allocate their cognitive capacity effectively towards essential and generative processing (i.e., coherence, signalling, redundancy, spatial and temporal contiguity principle), 2) managing essential processing, optimizing the process of building mental representations in learners' working memory (i.e., modality, segmenting, and pre-training principle), and 3) fostering generative processing, encouraging learners to consistently invest and maintain effort in comprehending the material (i.e., multimedia, personalisation, voice, and embodiment principle).

2.1.3 The role of emotions in learning

Both CLT and CTML have significantly advanced our understanding of how people learn from educational materials and provided foundational principles for designing effective instructional materials. However, with the literature on the topic expanding, one notable limitation of both theories emerged – while they both focus on cognitive processes, they largely overlook the impact of other facets of learning, such as affective, metacognitive, and social processes. Both theories have been criticized for not sufficiently incorporating the emotional aspects of learning by other researchers (Plass and Kalyuga, 2019) as emotions play a crucial role in cognitive processing, motivation, and memory, which are all vital elements of the learning process (Christenson et al., 2012).

Emotions, as fundamental components of human psychology, significantly influence behavior, decision-making, and learning. Emotions are complex states, triggered by various events or interactions which can positively or negatively affect individuals' psychological state. While there is no general agreement on the definitions of emotion, affect, and mood, the term affect can be used as an umbrella term and refers to a multifaceted phenomenon in which affective, cognitive, physiological, motivational, and expressive processes combine into an emotional episode (Pekrun and Linnenbrink-Garcia, 2014; Shuman and Scherer, 2014). Emotions are commonly described and understood through various models, including core affect frameworks proposed by Russell (1980) and Watson and Tellegen (1985).

The core affect model by Russell (1980) introduces a bidimensional space, representing emotions along two orthogonal, bipolar, and continuous dimensions: valence (pleasant-unpleasant) and arousal (activated-deactivated). According to Russell, all emotional experiences can be located within this circumplex structure, offering a simplified way to understand the spectrum of human emotions without categorizing them into discrete types (Russell, 1980). For example, enthusiasm would

be a pleasant and activating emotional state, and calmness a pleasant but deactivating state (Figure 2).

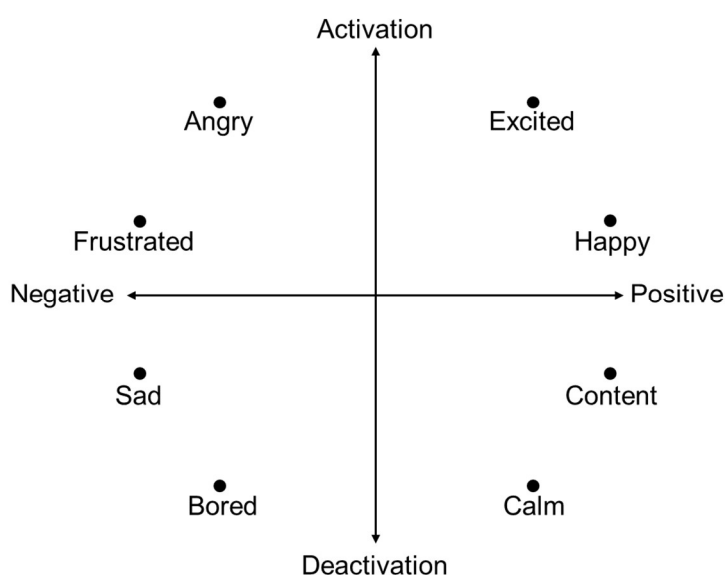


Figure 2: Model of core affect (Russell, 1980).

Watson and Tellegen (1985) proposed a similar two-dimensional affective space but emphasized positive affect and negative affect as independent dimensions, suggesting that individuals can experience high levels of both positive affect and negative affect simultaneously, unlike implied by the bipolar structure proposed by Russell. This model underscores the complexity of human emotions, suggesting that seemingly opposing emotional states can coexist (Watson and Tellegen, 1985). As can be seen in Figure 3, the Positive activation dimension ranges from pleasant states with high activation (e.g., enthusiastic, excited) to unpleasant states with low activation (e.g., drowsy, dull), while the other dimension, named Negative activation, spans from unpleasant high activation (e.g., nervous, distressed) to pleasant low activation states (e.g., relaxed, calm).

Due to the similarity between the two models, the authors later noted that the models are rotational variants of one another (Ekkekakis, 2013).

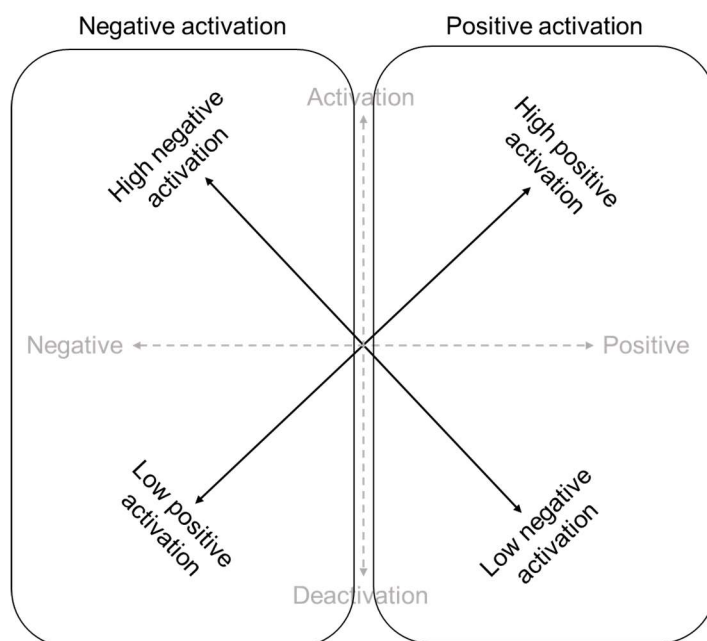


Figure 3: Watson and Tellegen's (1985) two-dimensional model (with Russell's model with grey dotted lines for comparison).

In the context of learning, emotional states can affect people's motivation and engagement as well as cognitive processes such as perception, memory, and problem-solving abilities. They can enhance or impair learning depending on their nature and intensity. They can either facilitate understanding and retention when positive or create barriers to learning when negative, playing a crucial role in the learning outcomes (Plass and Kalyuga, 2019).

The interplay between emotions, cognition, and learning suggests that instructional design strategies should consider not only the cognitive aspects of learning but also the emotional states of learners. Acknowledging and integrating the emotional dimension into the theories can lead to more comprehensive models of learning that better reflect the intricacies of human cognition and educational experiences. Therefore, research efforts in the last years have shifted to make a more comprehensive theory that also includes the latter factors.

2.1.4 Cognitive-Affective Theory of Learning with Media

The first theory to extend and modify the CTL and CTML by integrating motivational and emotional factors into the learning process was the Cognitive-Affective Theory of Learning with Media (CATLM) proposed by (Moreno, 2006) and Moreno and Mayer (2007). CATLM builds on the idea that learning with multimedia involves not only

cognitive processes but also affective, motivational, and metacognitive components, which interact to influence learning outcomes.

In other words, according to CATLM, learning is influenced by both the design of multimedia materials (e.g., text, images, animations) and the learner's emotional and motivational states. These states can affect how information is processed and retained. For example, pleasant emotions and motivation can improve cognitive processing, attention, and engagement, while unpleasant emotions can hinder learning by distracting the learner or reducing motivation (Moreno, 2006).

From the theoretical perspective, CATLM proposes additional assumptions to the three suggested by CTML (separate processing in dual channels, limited capacity, and active processing): 1) affective mediation, meaning that emotion and motivation mediate learning by enhancing or reducing cognitive engagement, 2) metacognitive mediation – the notion that self-regulation impacts learning by regulating cognitive processes and emotions, and 3) individual differences – the idea that variations in learners' prior knowledge and traits can affect the process of multimedia learning (Figure 4).

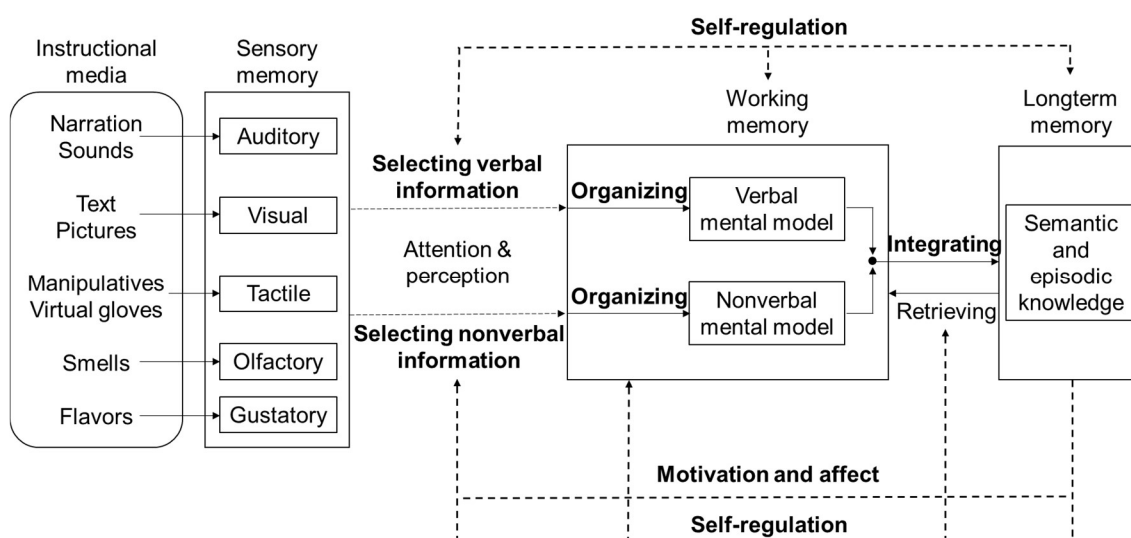


Figure 4: CATLM (Moreno, 2006).

The theory highlights the importance of designing educational media that not only presents information effectively but also addresses learners' emotional and motivational needs. The importance of adding emotional elements in multimedia learning, together with their role in reducing cognitive load and facilitating the integration of new information with existing knowledge, has been supported by various studies (Mayer, 2014; Park et al., 2014), leading to the concept of emotional design in multimedia learning, which

involves creating materials that elicit positive emotional responses to facilitate learning and retention.

2.1.5 Integrated Cognitive Affective Model of Learning with Multimedia

The Integrated Cognitive Affective Model of Learning with Multimedia (ICALM; Plass and Kaplan, 2016) represents another advanced conceptual framework that incorporates both cognitive and affective elements to better understand and optimize learning processes in multimedia environments. However, the ICALM proposes that affective processes such as emotions and motivation are not only mediators but are intricately intertwined with cognitive processes like memory and attention, making cognition and emotions inseparable. The authors stress the interaction between emotion and cognition, noting that as emotional processes place demands on cognitive resources, so do cognitive processes influence our emotional experiences. The implication for those designing educational materials is that the instructional design must not only focus on minimizing extraneous cognitive load but also prevent emotional load, while still ensuring that some degree of emotional arousal is maintained. The model thus proposes that effective multimedia learning involves not only the efficient processing of visual and auditory information but also the engagement of learners' emotions to enhance motivation and deepen understanding (Plass and Kalyuga, 2019).

2.1.6 Cognitive Affective Model of E-Learning

Recently, another model has been developed to specifically research the use of emotional elements in multimedia learning – the Cognitive Affective Model of E-Learning (Lawson et al., 2021b; Mayer, 2020). This model represents a framework built on all previously mentioned theories and research and is adapted to measure the effectiveness of using certain elements through the (onscreen) instructor in the e-learning materials in eliciting emotional responses in learners that in turn should affect their cognitive, behavioral, and learning outcomes.

The model consists of five steps: first, the emotional design intervention is introduced, such as the instructor displaying an emotional stance during the lesson. The intervention should be of such an intensity for the learners to notice it (for example, they would perceive and recognize the instructor's emotional stance). The third step of the model represents the learners experiencing the same emotion as it was displayed due to feeling a social connection with the instructor, which would lead to the learner exerting more or less (depending on the displayed emotion) effort into the learning (step 4). The

final step is the learning outcome, reflected in post-test performance. An example of this model from a study is shown in Figure 5.

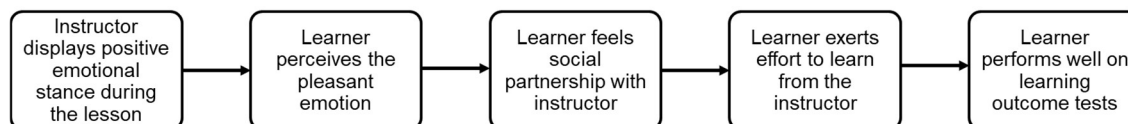


Figure 5: The Cognitive-Affective Model of E-Learning (Lawson et al., 2021b).

This model has been used in studies focusing on the effect of various types of instructors in multimedia educational materials, for example, onscreen human and virtual instructors (e.g., Horovitz and Mayer, 2021; Lawson et al., 2021c; Lawson and Mayer, 2021, 2022). However, many unanswered questions remain, such as how specific emotions expressed by the instructor affect those who learn with video and how to elicit those emotions in them. As research on online educational materials has been consistently showing, learning does not depend only on cognitive processes, which is why new research has to focus on components such as affective processing to advance our understanding on how to design effective multimedia learning materials (Mayer, 2024).

2.2 Emotional design principles

The approach that seeks to improve educational videos by manipulating affective-motivational factors is called emotional design. In the context of multimedia learning materials, emotional design features aim to impact learners' emotions that could promote learning (Plass and Kaplan, 2016). Emotions can substantially impact cognitive processes such as perception, attention, learning, and memory (Tyng et al., 2017). Furthermore, a meta-analysis on emotional designs in multimedia learning showed that integrating emotional designs can generally enhance learning outcomes, positive affect, and intrinsic motivation while reducing perceived difficulty (Wong and Adesope, 2020). However, their influence on learning may differ.

On one hand, they can improve learning by motivating learners to allocate more cognitive resources to the learning task by leveraging their interest and enjoyment, especially when learning for longer (Endres et al., 2020). For example, research has shown that experiencing positive emotions enhances motivation and learning (Um et al., 2012). On the other hand, emotional processing can increase cognitive processing,

imposing additional extraneous cognitive load and impeding learning (Plass and Kalyuga, 2019).

This area of research is still fairly fresh, with the first study specifically examining emotional design in multimedia learning conducted in 2012 (Um et al., 2012). This study demonstrated the importance of emotional factors in instructional design, as an emotional design intervention was shown to enhance learning by promoting positive emotions and cognitive processing, making it a promising field for further exploration.

2.2.1 Minimal manipulations and seductive details

Research on emotional design principles in multimedia learning can be categorized into two main branches: minimal manipulations and seductive details.

Minimal manipulations refer to subtle changes made to the design of learning materials to evoke positive emotions without altering the core instructional content. These manipulations aim to impact learner motivation and affect but maintain the educational content's integrity. For example, using round, human-like shapes and warm colours in educational materials can make the learning environment more pleasant and engaging, thereby inducing positive emotions and facilitating learning (Wong and Adesope, 2020). Studies by Mayer and Estrella (2014) have demonstrated that such minimal emotional designs can lead to better retention and understanding, indicating that even small changes in design can have significant impacts on learning efficiency (Mayer and Estrella, 2014).

Specifically, several studies have investigated the impact of an onscreen agent's emotional state on learners. In these studies, instructors conveyed their emotions through several social cues, such as facial expressions, gestures, body posture, vocal prosody, and anthropomorphic features (e.g., Lawson et al., 2021c, 2021a; Schneider et al., 2022; Um et al., 2012). Recent research demonstrates that learners can effectively recognize emotions exhibited by both human and virtual instructors (Horovitz and Mayer, 2021; Lawson et al., 2021b, 2021c) and that video instructors displaying pleasant emotions (e.g., being happy or content vs. displaying being frustrated or bored) aid learners in building a social connection during learning, pay more attention to the lecture, and score higher on a delayed (but not immediate) post-test (Lawson et al., 2021a), making a case for utilizing pleasant emotions in multimedia learning. These results laid evidence for the positivity principle, stating that people learn better from instructors who display pleasant or "positive" emotions compared to unpleasant emotions (Lawson et al., 2021a). However, the authors of these studies also highlighted the need for additional

research on the arousal/activity dimension of core affect in multimedia learning. Their findings indicated that participants had more difficulty distinguishing between active and passive instructors, suggesting that individuals are less attuned to the active-passive dimension of emotion compared to the pleasant-unpleasant dimension (Lawson et al., 2021a, 2021b, 2021c).

Seductive details, on the other hand, involve the addition of information that is interesting, but irrelevant to the instructional material itself (Harp and Mayer, 1997), intended to make the learning material more engaging. Examples of seductive details would include different types of decorative images, animations, fun facts or anecdotes, background music, sounds, etc. While these details can increase learner interest and engagement, they may also distract from the main instructional goals and lead to cognitive overload, thereby hindering the learning process. In fact, multiple meta-analyses reported mixed results, but highlighted the potential drawbacks of integrating seductive details into learning materials, showing that seductive details can indeed decrease overall learning performance (Rey, 2012; Sundararajan and Adesope, 2020).

Sung and Mayer (2012), for example, found that while graphics can enhance lesson enjoyment, they do not necessarily improve learning outcomes. While instructive graphics improved recall, seductive graphics hindered learning. Similarly, Park and others (2015) showed that seductive details boost positive emotions but hamper cognitive and learning performance. This effect is stronger for narrated seductive details compared to textual ones, as these can be ignored. A recent study confirmed the seductive details effect also during longer study sessions, affecting transfer but not retention (Bender et al., 2021). However, transfer was only impacted when participants were unaware of the irrelevance of the additional, seductive material. In other words, participants who were informed and recognized that certain parts of the study material were irrelevant did not experience a decline in performance due to seductive details. Finally, Schneider and others (2019) emphasized the importance of potential moderating variables, especially arousal. They found that the detrimental effect of seductive details on retention, transfer, and cognitive load was evident only when participants were in a lower state of arousal. When participants' activation levels were higher, the negative effects disappeared. Based on these findings, the authors caution against generalizing the negative effects of seductive details and recommend further research to explore the influence of arousal-enhancing features in multimedia learning environments.

2.2.2 Auditory emotional design in multimedia learning

In the realms of multimedia learning and emotional design, substantial emphasis has been placed on the visual components of instructional materials, such as text, graphics, colour, anthropomorphisms of non-human graphical elements, and animations. These elements have been the focus of numerous studies aiming to optimize instructional design principles to enhance learning outcomes. However, a vital but less scrutinized component of multimedia learning is the audio aspect, or what the viewers of the video can hear. There are four main types of auditory information that usually accompany visual content – narration, music, sound, and noise.

Despite its role in multimedia presentation, the auditory component has received considerably less attention, especially in educational research. In fact, two meta-analyses focusing on the impact of emotional design on learning included only studies (33 and 28 independent samples) with visual emotional design features, excluding any sound-related conditions (Brom et al., 2018; Wong and Adesope, 2020). A meta-analysis focusing only on seductive details (68 experiments), on the other hand, did include studies using auditory emotional design but did not provide any details regarding the type of auditory features used (Sundararajan and Adesope, 2020). The authors found a small to moderate negative effect ($g = -0.27$) on learning when seductive details were presented in audio form.

As multimedia technology evolves and becomes more sophisticated, the role of audio in learning environments cannot be overlooked. Similarly to visual cues, audio elements have the potential to reinforce learning, facilitate memory retention, and enhance learner engagement, particularly when effectively integrated with visual information, which applies both to verbal (narration) and nonverbal elements (emotional tone of the narration, sounds, music). However, despite the theoretical underpinnings supporting the role of audio in multimedia learning, empirical research focusing specifically on audio elements is limited. This gap highlights a need for comprehensive studies that investigate how different types of audio variables impact learning outcomes, and how audio interacts with the learning content to influence cognitive processes and affective states. The thesis will focus on two types of auditory emotional design – the emotional tone of the narration (as an example of an auditory minimal manipulation) and background music (representing an auditory seductive detail).

2.2.2.1 Emotional tone in narrator's voice in educational videos

As established, previous research has already examined the impact of onscreen instructors' emotions on learning, revealing not only the positive effects of pleasant versus unpleasant emotions (e.g., Lawson et al., 2021a, 2021b, 2021c) but also the differential impact of various pleasant emotions, such as enthusiasm versus calmness, on multimedia learning (Liew et al., 2017). Two competing hypotheses were tested: one positing that enthusiastic cues enhance positive emotions, thereby improving affective perceptions, intrinsic motivation, and cognitive outcomes; the other suggesting that additional emotions might increase extraneous cognitive load, negatively affecting outcomes. The findings favoured the former hypothesis, showing the beneficial effects of enthusiasm or activating emotions mediated by learners' positive emotions. However, these studies used onscreen pedagogical agents displaying multiple social cues, including facial expressions and body language (Lawson et al., 2021a, 2021b, 2021c). Since instructor presence in a video does not seem to have an effect on cognitive processing and learning in most learning domains (Beege et al., 2023; Heidig et al., 2024) and adding an additional pedagogical agent or just a video of the presenter can impose additional costs, time, and work for educators making the video, it makes sense to isolate the effect of voice alone on emotional, cognitive, and learning outcomes.

Narration refers to the spoken words in an instructional video and is used to explain the topic. It is the most commonly found auditory information type in educational videos, especially those that are more popular (ten Hove and van der Meij, 2015). On top of the verbal content presented through the spoken words, the human voice also conveys emotional information through nonverbal vocal expressions called prosody, referring to variations in pitch, loudness, rhythm, and voice quality in one's speech (Wilson and Wharton, 2006).

Research focusing solely on emotions conveyed through voice or narration remains limited. A recent study found that learners can discern emotional tone in voice just as accurately as with an onscreen instructor present who offers additional social cues like eye gaze, gestures, facial expression, and body stance (Lawson and Mayer, 2021), giving further credit to researching the isolated effect of social cues conveyed solely through voice. However, as in similar studies, the authors of the study also noted that while participants could effectively distinguish between positive and negative emotions, they struggled more with differentiating between emotions of the same valence.

In two experiments comparing an enthusiastic and calm narrator, researchers found that participants who watched the video with the enthusiastic narrator (who used

significant changes in tone and pitch; Collins, 1978) viewed the instructor more positively and performed better on knowledge tests, whereas participants who watched the video with a calm narrator (pleasant, calm tone, minimal pitch variation) reported experiencing higher germane cognitive load (Liew et al., 2020).

A similar study (Beege et al., 2020) used a 2x2 factorial design with participants assigned to one of four conditions: high vs. low mental load and enthusiastic vs. neutral pedagogical agent voice. Participants underwent a learning session followed by a multiple-choice test to measure learning outcomes. Results were mixed: with low mental load, the agent's enthusiastic voice improved performance on multiple-choice tests but not on open-ended questions. In contrast, under high mental load, participants scored higher on multiple-choice tests with a neutral voice, showing that enthusiasm may have both a beneficial and hindering effect on learning.

Research on the effect of emotional tone in the narrator's voice is thus not only limited but also mixed. Providing additional emotional cues in the voice may improve students' outcomes (Liew et al., 2020), but may also make the instructional message more complex, requiring more information processing and eventually impeding learning (Beege et al., 2020), especially when learning from a video in one's non-native language (Vanlancker-Sidtis, 2003). A between-subjects experiment compared non-native English speakers watching videos in English narrated by a strong-prosodic human voice, a weak-prosodic human voice, or a modern computer voice (Davis et al., 2019). There were no significant differences in cognitive load, retention, or agent persona between the two human voices, though some differences were noted compared to the computer voice, painting an even more complex picture of the isolated effect of the instructor's narration. Despite focusing on non-native English speakers, the authors emphasized that the learning episode was brief (between 3 and 3.5 minutes) and the sample, consisting of English majors or double majors, may not have been representative of the broader non-native speaker population. Therefore, they stressed the importance of further investigating the impact of different modalities of human voice on a more diverse non-native sample.

The potential of narration to convey emotions effectively, even without visual cues, has significant implications for the design of educational videos. This approach can reduce production costs and time while still providing an engaging learning experience. However, it is essential to consider the balance between providing emotional engagement and avoiding cognitive overload, especially when targeting learners who will view the videos in their non-native language, so additional testing is needed.

2.2.2.2 Music in educational videos

While changing the emotional tone of the narrator's voice provides an example of a minimal manipulation, adding background music to a video exemplifies the addition of seductive details. Background music is added in roughly 60% of educational videos on YouTube, with two-thirds of them playing the song throughout the whole video and a third of videos combining both music and narration. Additionally, instructional videos with background music seem to be more popular than those without (ten Hove and van der Meij, 2015), so it makes sense to explore the intricacies of incorporating music in such a context.

Some of the most important acoustical features in music that affect how listeners perceive the emotions expressed in the music are tempo (number of beats per minute – fast or slow) and mode (specific set of pitches/notes used in a musical segment – minor or major) (Gagnon and Peretz, 2003; Juslin and Laukka, 2003), with tempo being the more prominent. Musical pieces composed in a major mode are usually perceived as happy, while those in a minor mode are typically seen as sad. Additionally, a faster tempo is associated with more arousing emotions (e.g., happiness, anger), while a slower tempo is linked to less arousing emotions (e.g., calmness, sadness; Ho and Loo, 2023).

Music induces emotions in listeners through various mechanisms, one of which is emotional contagion. Emotional contagion in music refers to the phenomenon where music with a specific emotional expression can evoke the same emotion in the listener with the same or lower intensity (Juslin and Västfjäll, 2008; Schubert, 2013). For example, if a piece of music conveys happiness, the listener may feel happier; and if the music conveys sadness, the listener may feel sad. A study investigating emotional and physiological responses to different positive valence music pieces had participants listen to two songs chosen by the researchers—one low-arousal and one high-arousal—as well as a song self-selected by the participants and described as "uplifting." The findings revealed that the self-chosen uplifting song generated the most joy, the low-arousal piece was linked to the highest relaxation and lowest anxiety levels, while the high-arousal jazz piece produced a much smaller mood enhancement. The self-selected uplifting song also resulted in the highest physiological activity, whereas both researcher-selected pieces, regardless of arousal level, had a similar impact on participants' physiological activity (Lynar et al., 2017). Another finding of the study is that participants who were experiencing high levels of psychological distress saw the most significant improvements in their emotional state from listening to music.

Similarly to the emotional tone of the voice, music is therefore added to the video with the goal of affecting mood and arousal (Salimpoor et al., 2009), which in turn affects attentional resources, cognitive performance, and learning outcomes (Husain et al., 2002). Arousal or activation increases learning up to a point, but too much of it decreases students' performance (Teigen, 1994). This principle, known as the Yerkes-Dodson law, is illustrated as an inverted U-shaped curve showing the relationship between learning and arousal levels (Figure 6). Arousing stimuli can serve as alert signals, capturing attention and prompting quick responses, while also being better remembered (Chung et al., 2015; Schneider et al., 2019). In other words, there is likely an optimal level of arousal that is not too low (resulting in no activation) and not too high (causing avoidance behavior), but at a level that effectively engages learners.

Embedding music in an educational video seems a simple way to increase the learners' activation and their willingness to engage with the content, especially songs with a higher tempo (Husain et al., 2002). On the other hand, however, music as an irrelevant, seductive detail, also poses an additional load, which can impact learning negatively, so the final effect is not clear.

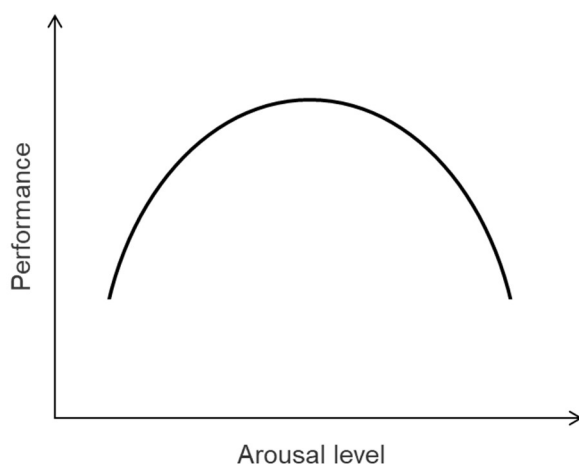


Figure 6: The Yerkes-Dodson law (Teigen, 1994)

Meta-analyses report mixed and inconsistent results regarding the effects of background music on cognitive performance and learning, with studies showing positive, neutral, and negative outcomes (de la Mora Velasco and Hirumi, 2020; Kämpfe et al., 2010). There are also several methodological issues with the research on the effect of background music on learning, making it harder to generalize the findings. For example, a seminal study with two experiments on the effects of sounds in multimedia learning found a hindering effect of adding music to the presentation, but the instrumental clip they used was 20 seconds long and played in a loop (Moreno and Mayer, 2000).

The authors of the meta-analysis highlight a great variation in background music interventions, task type and difficulty, and learners' characteristics, and emphasize the need for detailed reporting of music stimuli elements, such as genre, activation level, and valence, which are often omitted (de la Mora Velasco and Hirumi, 2020). In addition, studies that do report the genre most frequently use classical music (e.g., Lehmann, Hamm, and Seufert, 2019) or music with human vocals, overlooking the type of songs that are most commonly used in contemporary educational videos (e.g., ambiental music). Including information such as type of music and music tempo is vital, as these can greatly differ in their effect on task performance. For example, music with a high tempo was found to be more detrimental to reading comprehension and free recall than music with a slow-tempo in one's native language as it includes a higher number of auditory events per unit of time, which consumes more of the listeners' limited attentional resources, while the slow tempo music allows for better recovery from acoustic interference (Cassidy and MacDonald, 2007; Thompson et al., 2011). However, this effect reverses in foreign language learning, where fast-tempo music enhances performance more than slow tempo music (Su et al., 2023). The impact of tempo (and music in general) also varies with task difficulty, affecting easier tasks more than difficult ones (Meyerhoff et al., 2022; Su et al., 2023).

Another important factor is whether the music includes human vocals. For instance, Alley and Greene (2008) examined the effects of vocal music, instrumental music, irrelevant speech, and silence on a working memory task. They discovered that both irrelevant speech and vocal music disrupted working memory, whereas participants who listened to instrumental music performed similarly to those in the silent group, implicating that purely instrumental songs would work better as background music in instructional videos.

Some studies found that background music, either low- or high-arousal, fails to elicit an emotional arousal response in participants and does not have either a decremental or incremental effect on learning or task performance (Du et al., 2020; Jäncke and Sandmann, 2010; Lehmann and Seufert, 2017).

Adding to the complexity, most studies do not include background music embedded within a multimedia presentation. In fact, in the latest meta-analysis, only three studies examined music embedded in multimedia, reporting positive effects on motivation, recall, and language learning (de la Mora Velasco and Hirumi, 2020). On the other hand, a recent study found no effect of background music on recall, comprehension, and extraneous cognitive load, but a beneficial effect for germane cognitive load and transfer,

meaning that the music aided in engaging more intensively in more complex tasks (Lehmann et al., 2019).

Researchers are also exploring whether individual differences, especially personality traits, influence the effects of background music on cognitive processes and learning. According to Eysenck's (1967) arousal theory, introverts, who tend to have higher cortical activity and arousal, avoid additional stimulation, while extraverts seek it out. Thus, it has been hypothesized that background music in multimedia learning environments might have different effects depending on the learner's level of extraversion, in particular, that introverts would find background music more distracting than extraverts (Cassidy and MacDonald, 2007) and that it would negatively impact introverts more. However, research results have been mixed (Cassidy and MacDonald, 2007; Dobbs et al., 2011; Lehmann et al., 2019).

Generally, the research on background music in educational contexts and multimedia learning presents a mixed picture. While background music has the potential to enhance the learning experience by increasing engagement and motivation, its impact on cognitive learning outcomes is less clear and might be context-dependent. The selection of background music should be carefully considered and researched to ensure it supports rather than detracts from the educational objectives.

2.3 Learning in a foreign language with same-language subtitles

Another research gap in the scientific literature on the role of auditory emotional design in multimedia learning is that most studies focus solely on instructional materials in the learners' native language. With evidence of differences in learning processes when using multimedia presentations in one's native language versus a foreign language (Davis and Vincent, 2019; Lee and Mayer, 2018; Mayer and Fiorella, 2014) or even in a different dialect (Rey and Steib, 2013; Schneider et al., 2015), it is important to expand the research and include a variety of participants and contexts.

In a globalized world, a lot of online multimedia content, including educational resources, is available in English. However, many people who consume English learning materials are non-native speakers, and learning in a foreign language demands additional mental resources, which can overwhelm the learner's cognitive system, potentially hindering the learning process (Sweller et al., 1998). This is particularly evident in online multimedia learning environments where verbal and visual information are processed simultaneously.

Narrated words are transient, meaning that people, especially those with lower language proficiency and who have not yet automated their phonological processing of sounds in the foreign language, may have trouble segmenting the continuous flow of sounds into discreet words and adequately processing the spoken information, as they are consciously trying to perceive each word (Leahy and Sweller, 2011; Mayer et al., 2014). Written words, on the other hand, are available for longer and allow learners to revisit them (Mayer et al., 2020). Providing subtitles is therefore an easy and relatively inexpensive way of providing language information through the visual channel, complementing the auditory channel. Since educational materials are open to people all around the world, and providing subtitles in all languages is impractical, subtitles in the video's language or same-language subtitles (SLS) are a great alternative to subtitles in the learner's first language that ensures the video's accessibility to a wider audience. On one hand, SLS can maintain longer word availability (compared to transient narrated words), facilitating word encoding and helping learners with deeper processing of the content. On the other, the written text in addition to narration may be redundant and can compete for finite cognitive resources that are needed during learning as the SLS makes them split their visual attention between the video and SLS simultaneously, creating additional extraneous cognitive load. For example, a study using eye-tracking found that there is an approximately 2-second delay before viewers of videos shift they focus from subtitles to newly appearing graphics (Persson et al., 2019), indicating the need to further investigate all the effects of adding SLS to video.

A recent literature review summarizes studies on the effect of adding subtitles into three research domains: 1) using subtitles when learning content in one's native language, 2) using subtitles when learning a foreign language (with subtitles either in one's native or the foreign language that is to be learned – SLS), and 3) using subtitles when learning content in a foreign language (again, the subtitles being either in one's native or foreign language – SLS), with the last one having the least amount of research evidence (Pannatier and Betrancourt, 2019).

First, for people learning in their native language, SLS have been shown to have no or even a harmful effect on learning (Lebeničnik et al., 2020; Mayer and Fiorella, 2014) due to the modality principle, suggesting that people learn more effectively from graphics accompanied by spoken words rather than graphics paired with written words, and the redundancy principle, indicating that learning is improved when graphics are combined with spoken text alone, rather than with both spoken and written text (Mayer, 2014). Second, when learning in one's foreign language, most research focused on the effect of SLS when learning the language in question, especially listening comprehension and

vocabulary learning, yielding positive results (Perez, Van Den Noortgate, and Desmet, 2013). Finally, research on learning a non-language related subject in one's non-native language is not so clear.

Different studies show mixed results. An experiment with 374 Korean college students who watched a 16-minute English video about Antarctica revealed that those in the group with added SLS scored higher on a comprehension test and reported significantly lower difficulty and effort in learning compared to those without SLS (Lee and Mayer, 2018). Similarly, a study involving 73 undergraduates from two Taiwanese universities learning about brain anatomy and cognitive functions found that, after accounting for prior knowledge and English proficiency, students who watched the video with SLS performed better on a post-test and experienced lower cognitive load than their peers without SLS (Lin et al., 2016). These findings indicate that when learning in a foreign language, the redundancy and modality principles do not apply as they do in one's native language, and the split-attention effect between the two visual sources is not observed.

In contrast, other studies found null results, meaning that SLS did not have either a beneficial nor a detrimental effect on retention, transfer, cognitive load, self-reported enjoyment, and perceived difficulty of the lesson (Liu et al., 2018; Matthew, 2020; Mayer et al., 2014; Pannatier and Bétrancourt, 2024; van der Zee et al., 2017). For example, a recent study failed to find any effect of either SLS or subtitles in the participants' own language on learning performance, cognitive load, and situational interest, regardless of the level of proficiency in the language of the video (Pannatier and Bétrancourt, 2024). However, even one of these studies still found that learners who viewed the lesson with SLS reported exerting significantly less effort in understanding the lesson compared to the no-SLS group, although this perception did not translate into better test results (Mayer et al., 2014).

Despite the conflicting findings, an overview of several studies (Gernsbacher, 2015) shows that SLS have the potential to benefit many viewers, not just those learning a new language or with hearing impairments. At least in some contexts, captions may be able to improve comprehension, attention, and memory for the video content, suggesting that SLS might enhance the educational value of different video materials. Further research is needed to confirm whether SLS are generally beneficial for learning purposes, and which characteristics of the learning context make SLS more or less useful.

Non-native learners using English resources may experience varying outcomes based on their English language proficiency (Lin et al., 2016). The interaction between English language knowledge and learning from English multimedia resources, such as

those with subtitles, demands further examination. Research should also explore how different levels of language proficiency impact comprehension and retention when learning from English-language materials.

3 EMPIRICAL PART

3.1 Research problem, purpose, and objectives

Until now, research on multimedia learning has mostly focused only on cognitive factors, and research on the principles of emotional design only on interventions related to the visual channel. The research problem we currently face is therefore the lack of knowledge about how learning can be affected by audio stimuli designed to change students' emotions. Sound is an important part of educational videos, which is also worth exploring in the context of emotional design. Given that prior research has demonstrated the beneficial impact of pleasant emotions on learning, it makes sense to further research these in particular, for example examining their level of activation, as it can affect performance (Lawson et al., 2021a, 2021b, 2021c; Teigen, 1994). An additional research problem is also the fact that most studies used learning materials in the students' native language, which does not reflect modern online learning, which is mostly conducted in English. The university population has a relatively good knowledge of English and therefore increasingly uses English videos in both their formal and nonformal learning. Our materials will thus be in English, which will not only allow for greater comparability with international studies, but also to test whether the addition of same language subtitles is beneficial.

The purpose of the dissertation is to explore how auditory emotional design and same language subtitles impact learning of students who watch a multimedia lesson about wood as a construction material in a foreign language. We will gain this knowledge through videos on the topic of sustainable construction – a topic that is unknown to most people despite its significant contribution to the fight against climate change.

The aim of the dissertation is to conduct two experiments to determine the effect of the narrator's emotional tone expressed only through voice, the addition of background music with different levels of activation, and SLS, as reflected in the learners' learning, cognitive load, and affective variables. Since they are the most likely to be used in instructional videos in practice, we will focus only on pleasant emotions in videos,

specifically on one activating (enthusiasm) and one deactivating (calmness) pleasant emotion, that will be expressed either through the narrator's voice or background music.

Specifically, Study 1 will have the following objectives:

- 1) To determine whether the emotional tone of a disembodied instructor (narrator) affects the learning, cognitive, and affective variables of participants' who are watching learning videos in their non-native language;
- 2) To investigate the influence of SLS on learning from the videos;
- 3) To analyse potential variations in results based on participants' English proficiency.

In turn, the following objectives were set for Study 2:

- 1) To examine the effect of embedded background music in educational videos on learners' cognitive, affective, and learning metrics;
- 2) To assess whether these effects differ depending on whether the background music is lively and calm;
- 3) To consider the potential influence of individual differences, such as English proficiency and the relevance of the study program to the content of the educational videos, on the results.

In summary, research on emotional design and SLS provide two competing theories and findings; while additional cues may increase engagement and aid learners in processing information in a non-native language, they may also increase extraneous cognitive load and detract from learning. For this reason, no predictions on the direction of effects will be made, instead focusing on the differences. In general, we predict that learners will distinguish between activating and deactivating positive emotions expressed either through voice or music and that results will differ based on the activation level of the emotion and the presence or absence of SLS.

3.2 Pre-study 1: Recognizing human emotion from the narrator's voice

A preliminary study was conducted to determine if there was a discernible difference between the enthusiastic and calm voice prosody of the narrator in the videos and if the recorded material could be used as an independent variable.

3.2.1 Research hypotheses

Three preliminary hypotheses were made:

Preliminary Hypothesis 1: Enthusiastic videos will be rated significantly higher on the enthusiastic scale than the calm videos.

Preliminary Hypothesis 2: Calm videos will be rated significantly higher on the calm scale than the enthusiastic videos.

Preliminary Hypothesis 3: Enthusiastic videos will be rated significantly higher on the activation scale than the calm videos.

3.2.2 Methodology

3.2.2.1 Research design

Pre-study 1 was conducted as an online experiment with a within-subjects design where participants viewed and rated short clips taken from the instructional material. All respondents watched two sets of video clips – five clips with a calm voice and five identical clips but with an enthusiastic voice. Participants viewed the video clips in a randomized order and rated the valence, activation level, and the narrator's expressed emotion on a Likert-type scale. These types of scales were used in previous studies (Lawson et al., 2021c; Lawson and Mayer, 2021).

3.2.2.2 Participants

A convenience sample of 209 respondents aged from 16 to 72 ($M_{age} = 30.43$, $SD_{age} = 11.12$) participated in the study, with 132 identifying as female, 69 as male, three as non-binary, and five declining to disclose their gender. 49 participants were originally from Slovenia while the rest was from different countries around the world (Table 1). 47 participants filled the questionnaire in Slovene while the rest filled it in English. More than three quarters of participants had at least a bachelor's degree (78.95%) and the majority was either a university student or an employee (88.04%; Table 2). Respondents had a relatively high subjective English listening ability ($M = 6.14$, $SD = 1.08$; non-native English

speakers only: $M = 5.88$, $SD = 1.09$), low prior knowledge about the topic of the videos ($M = 2.90$, $SD = 1.64$), and moderate interest in the topic ($M = 3.73$, $SD = 1.91$).

Table 1: Country of participants ($n = 209$)

Country	n	$f\%$
Slovenia	49	23.44%
United Kingdom	27	12.92%
England	18	8.61%
United States of America, undisclosed	16	7.66%
Poland	12	5.74%
Germany	11	5.26%
Italy	6	2.87%
Brazil	5	2.39%
Australia	4	1.91%
Finland, Latvia, Norway, Scotland	3	1.44%
China, Hungary, India, Ireland, Malaysia, Russia, South Africa, Suriname	2	0.96%
Belgium, Bulgaria, Canada, Croatia, Egypt, France, Ghana, Israel, Jamaica, Malta, Netherlands, Spain, Sweden/New Zealand, Taiwan, Turkey, Wales, Zambia	17	0.48%

Note. Countries with the same number of participants have been grouped together

Table 2: Educational level and status of participants ($n = 209$)

Education	n	$f\%$
Primary education	3	1.44%
Secondary education	37	17.70%
Bachelor's degree (first Bologna cycle or equivalent)	78	37.32%
Master's degree (second Bologna cycle or equivalent)	55	26.32%
Doctorate degree or equivalent	32	15.31%
Undisclosed	4	1.91%
Status		
High school student	6	2.87%

University student	95	45.45%
Employed	89	42.58%
Unemployed	8	3.83%
Retired	1	0.48%
Employed and student	4	1.91%
Self-employed	1	0.48%
Undisclosed	5	2.39%

3.2.2.3 Materials

The materials were 10 videoclips ranging from 33 to 65 seconds in duration (see Appendix 1). These clips were taken from the video presentation on wood as a building material that were to be used in Study 1 and 2 and are better described in section 3.3.2.3. The narrated PowerPoint presentations were recorded by a woman with a Standard American English accent reading a script. Five clips were portrayed with an enthusiastic voice prosody and the other five in a calm voice. The narrator was instructed to portray the emotions in a realistic and nonexaggerated way. Several versions of the recordings were made based on feedback from the candidate. For the enthusiastic clips, the narrator was instructed to use an uplifting intonation and make regular changes in tone and pitch, while for the calm version, the narrator held the tone and pitch relatively constant (Collins, 1978).

3.2.2.4 Instruments

Two versions of the survey were made – a Slovene and an English version. Both versions contained the same videoclips in English, the only difference was the language of the rating scales and questions.

First, respondents were asked to rate their prior knowledge and interest in the topic of wood as a building material on a 7-point scale (1 – very low/not interested, 4 – moderate/neither not interested, neither interested, 7 – very high/interested).

Then, participants rated each videoclip on seven items, which were adapted from similar studies (e.g., Lawson and Mayer, 2021; Lawson et al., 2021). The participants were asked to rate the extent to which they thought the narrator expressed five emotions: enthusiastic, calm, frustrated, happy, and bored on a 7-point rating scale (1 – not at all, 4 – somewhere in between, 7 – extremely). The item “happy” was added as a pleasant emotion with an activation level between calmness and enthusiasm and the items with

“frustration” and “boredom” were added to provide the participants with examples of unpleasant emotions with high and low activation level. Rating of different items was chosen as this type of rating scale provides more information than a forced choice rating between calm and enthusiastic.

After rating the emotions, the participants rated the activation level and pleasantness of the narrator video on a 9-point scale ranging from extremely passive/unpleasant to extremely active/pleasant.

In the end, participants rated their English listening ability on a 7-point scale from very low to very high and provided demographic information regarding their country of origin, gender, education, and education/employment status.

3.2.2.5 Data collection

People over the age of 15 were invited to participate over the candidate's and InnoRenew CoE's social media. The survey was displayed on the online platform 1ka.si (Faculty of Social Sciences, University of Ljubljana, 2022). Before watching the clips, there was a ten second video prompting the participants to adjust the volume settings to ensure they hear the spoken text clearly. The videos (together with their rating scales) were shown in a randomized order.

Data collection lasted from December 2021 to February 2022. Respondents received no incentives for participation.

3.2.2.6 Data analysis

Data was analysed using the open-source software R (R Core Team, 2020) and jamovi (The jamovi project, 2022). Individual ratings of five enthusiastic clips and five calm clips were averaged (arithmetic mean) to create one (averaged) rating for the enthusiastic videos and one for the calm videos. Assumptions of sphericity and normality were checked by conducting the Mauchly's test of sphericity and visually inspecting Q-Q plots, respectively. Averaged ratings were then compared by using repeated measures ANOVAs with a Greenhouse–Geisser correction for lack of sphericity and post-hoc pairwise *t*-tests with a Bonferroni correction.

3.2.3 Results and interpretation

Mauchly's tests were performed to check the assumption of sphericity predicting that the variances of the differences between several conditions are equal. Both in the case

of enthusiastic ($W = 0.172$, $p < .001$) and calm videos ($W = 0.289$, $p < .001$, $\varepsilon = 0.709$), the assumption of sphericity was violated, so further comparisons were made with a Greenhouse–Geisser correction. The Greenhouse–Geisser correction was chosen instead of the Huynh–Feldt correction due to the fact that the Greenhouse–Geisser value ε was smaller than the rule of thumb number 0.75, in which case the Greenhouse–Geisser correction is recommended (Field, 2018; Navarro and Foxcroft, 2022).

Averaged ratings of the enthusiastic videos can be seen in Figure 7. An ANOVA on the averaged ratings of the enthusiastic videos was conducted, finding a significant main effect, $F(2.25, 468.60) = 613.00$, $p < .001$, $\eta^2_p = 0.75$. Pairwise comparisons between the enthusiastic rating ($M = 5.08$, $SD = 0.97$) and ratings on other emotional items shown in Table 3 reveal that the enthusiastic videos were perceived as significantly more enthusiastic than all other emotions, including calmness ($M = 4.37$, $SD = 0.98$), confirming Preliminary Hypothesis 1.

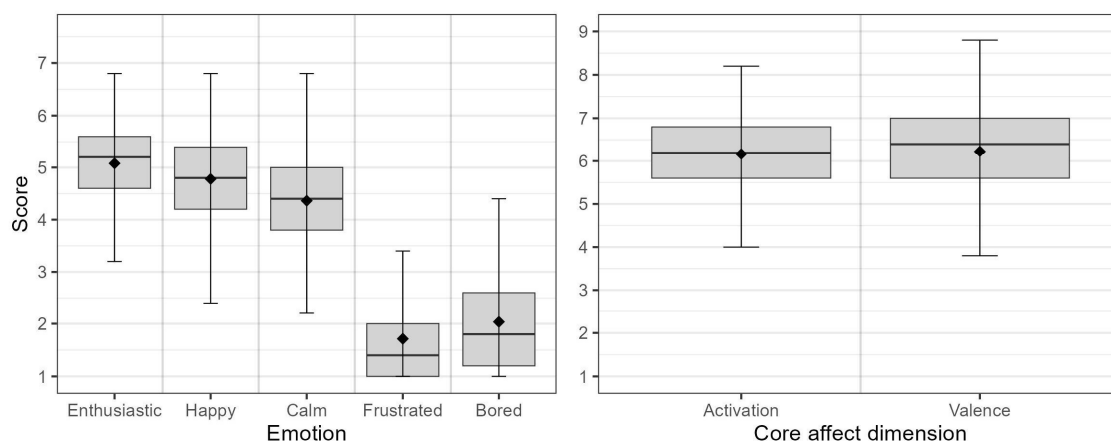


Figure 7: Averaged ratings of the enthusiastic videos.

Table 3: Paired t -tests comparing enthusiastic and calm ratings in enthusiastic and calm videos with other emotions

	t	p	Mean difference [95% CI]	d [95% CI]
Enthusiastic				
Calm	8.03	< .001	0.71 [0.54–0.88]	0.56 [0.41–0.70]
Happy	7.04	< .001	0.30 [0.21–0.38]	0.49 [0.34–0.63]
Frustrated	33.78	< .001	3.37 [3.17–3.57]	2.34 [2.07–2.60]
Bored	27.68	< .001	3.05 [2.83–3.27]	1.91 [1.69–2.14]
Calm				
Enthusiastic	33.21	< .001	3.16 [2.97–3.35]	2.30 [2.04–2.56]

Happy	31.81	< .001	3.05 [2.86–3.24]	2.20 [1.95–2.45]
Frustrated	21.24	< .001	2.73 [2.47–2.98]	1.47 [1.27–1.66]
Bored	4.93	< .001	0.58 [0.35–0.81]	0.34 [0.20–0.48]

Note. $df = 208$

A significant main effect was also found in an ANOVA on the averaged ratings of calm videos (see Figure 8), $F(2.89, 601.34) = 378.41$, $p < .001$, $\eta_p^2 = 0.65$. Paired t -tests (Table 3) showed that participants rated calm videos as significantly higher on the calm item ($M = 5.33$, $SD = 0.98$) compared to the enthusiastic ($M = 2.17$, $SD = 0.88$) and all other discrete emotion items, confirming Preliminary Hypothesis 2.

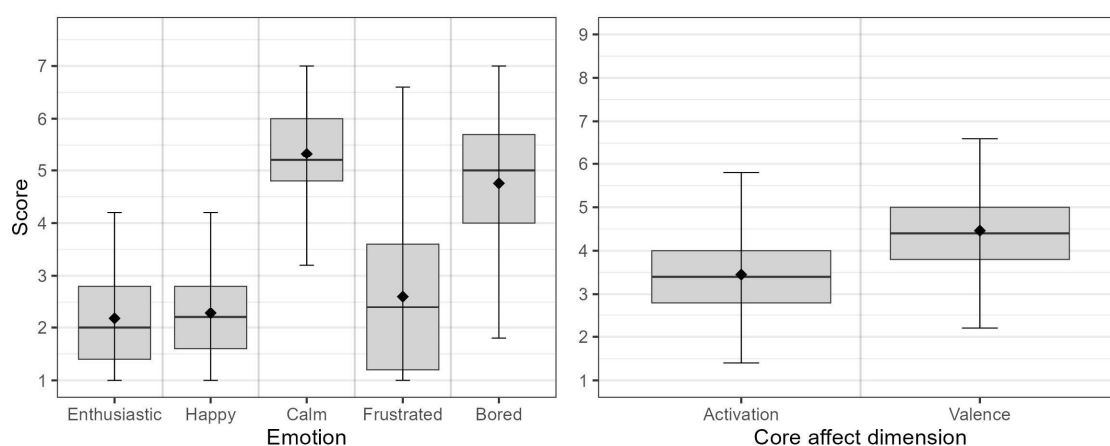


Figure 8: Averaged ratings of the calm videos.

We also compared the enthusiastic and calm videos by ratings of the two core affect dimensions – activation level and valence. The enthusiastic videos ($M = 6.17$, $SD = 0.94$) were found to have a significantly higher activation level compared to the calm videos ($M = 3.45$, $SD = 0.99$; $t(208) = 30.62$, $p < .001$, mean difference = 2.72, 95% CI [2.55 – 2.90], $d = 2.12$, 95% CI [1.87 – 2.36]), giving support to Preliminary Hypothesis 3. However, enthusiastic videos ($M = 6.23$, $SD = 0.96$) were also perceived to be significantly more pleasant than the calm videos ($M = 4.46$, $SD = 1.02$; $t(208) = 21.89$, $p < .001$, mean difference = 1.76, 95% CI [1.60 – 1.92], $d = 1.51$, 95% CI [1.31 – 1.71]).

Summarized, the results of Pre-study 1 indicate that the voice prosody in each type of video corresponds to the planned emotion of the narration and that there are significant differences between the enthusiastic and calm videos. This means that the videos can be further used as an independent variable in Study 1. Ideally, the enthusiastic and calm videos would differ only in terms of activation level and not valence. However, the difference in pleasantness is much smaller than the one in activation level, so we decided to use the videos as they were in the following experiment.

3.3 Study 1: Experiment on the effect of emotional tone in the narrator's voice and same-language subtitles

3.3.1 Research hypotheses

Eight hypotheses were made for Study 1:

H1: Participants with an enthusiastic narrator will rate the videos significantly higher on the enthusiastic and activation scales, while participants with a calm narrator will rate the videos significantly higher on the calm scale.

H2: Participants with an enthusiastic narrator will develop significantly different feelings of social partnership with the narrator than participants with a calm narrator.

H3: Participants with an enthusiastic narrator will have higher levels of positive activating emotions than participants with a calm narrator.

H4: Participants with an enthusiastic narrator will have higher situational interest than participants with a calm narrator.

H5: Participants with an enthusiastic narrator will have significantly different levels of cognitive load than participants with a calm narrator.

H6: Participants with an enthusiastic narrator will have significantly different learning outcomes than participants with a calm narrator.

H7: Participants with SLS will have significantly different levels of cognitive load than participants without them.

H8: Participants with SLS will have significantly different learning outcomes than participants without them.

3.3.2 Methodology

3.3.2.1 Research design

Study 1 was conducted as a quantitative experiment with a 2x2 between-subjects design and we used descriptive and causal experimental methods. The first factor was emotional tone of the narration in the videos – one group watched learning videos being narrated with an enthusiastic voice, while the other group watched videos narrated with a calming voice. The second factor in the experiment was inclusion of SLS – one group learned from videos without subtitles and the other group learned from videos that had SLS embedded. Taken together, four experimental conditions emerged: videos with a calm narrator and without added SLS (group C), videos with a calm narrator and with SLS added (C+S), videos with an enthusiastic narrator and without added SLS (group

E), and videos with an enthusiastic narrator and SLS added to the videos (group E+S). Participants were randomly assigned to each group. All participants went through the same experimental procedure, the only difference being the type of videos they were watching. Seven days after the experiment, participants were invited to participate in a second part of the study that included answering the same set of questions.

3.3.2.2 Participants

229 students participated in Study 1, but data from three participants were omitted from further analysis due to their failure to adhere to the study protocol and unreliable responses (providing only extreme or middle-range data through the whole survey), lowering the total number of respondents to 226. 81.42% of participants were students from various faculties from University of Primorska (UP) in Slovenia and 18.58% were students from the Norwegian University of Life Sciences (NMBU). In general, participants from UP were students from social science study programs while students from NMBU came from study programs in the life sciences. The majority of the participants were enrolled in a bachelor's degree program, while three reported being at the master's level. Three participants selected the "other" option but did not provide further details. The average age of the sample was 20.39 years ($Mdn = 20$, $SD = 2.65$) with the youngest participant being 18 years old and the oldest 45 years old. 183 were women, 41 men and two participants did not disclose their gender. The average age of the Slovenian sample was 19.80 ($Mdn = 19$, $SD = 1.51$) and the average age of their Norwegian counterparts was 23.00 ($Mdn = 22$, $SD = 4.46$).

Table 4 reports demographic statistics (gender, study program, study year, and country of residence) of the sample in total and divided by experimental condition/group. Two participants terminated their participation prematurely, so partial data will be analysed in their case.

Table 4: Demographics divided by country and in total

	Slovenia (N = 184)	Norway (N = 42)	Total (N = 226)
	n (%)	n (%)	n (%)
Gender			
Female	160 (70.80%)	23 (10.18%)	183 (80.97%)
Male	23 (10.18%)	18 (7.96%)	41 (18.14%)
Undisclosed	1 (0.44%)	1 (0.44%)	2 (0.88%)
Study program			
Management (UP)	58 (25.66%)		58 (25.66%)
Pedagogy (UP)	24 (10.62%)		24 (10.62%)
Pre-school Teaching (UP)	39 (17.26%)		39 (17.26%)
Primary School Teaching (UP)	63 (27.88%)		63 (27.88%)
Ecology and nature management (NMBU)		17 (7.52%)	17 (7.52%)
Forestry (NMBU)		16 (7.08%)	16 (7.08%)
Geomatics (NMBU)		1 (0.44%)	1 (0.44%)
Landscape architecture (NMBU)		2 (0.88%)	2 (0.88%)
Property development		4 (1.77%)	4 (1.77%)
Renewable energy (NMBU)		2 (0.88%)	2 (0.88%)
Study year			
1st (bachelor's)	142 (62.83%)	15 (6.64%)	157 (69.47%)
2nd (bachelor's)	27 (11.95%)	16 (7.08%)	44 (19.47%)
3rd (bachelor's)	15 (6.64%)	5 (2.21%)	21 (9.29%)
4th (or 1st master's)		2 (0.88%)	2 (0.88%)
5th (or 2nd master's)		1 (0.44%)	1 (0.44%)
Other		3 (1.33%)	3 (1.33%)

94 or 41.59% of those students participated also in the second part of the study (3 from Norway and 91 from Slovenia). Specifically, there were 22 students from Group C, 25 students from Group C+S, 26 students from Group E, and 21 students from Group E+S that participated in the delayed testing.

3.3.2.3 Materials

The independent variables were introduced via learning videos in English that were made specifically for these studies as a combination of slides and narration. Five different videos were made with a combined duration of 24 minutes. We chose to make the videos this length because research suggests that the impact of video design principles tends to be greater with longer video durations (de la Mora Velasco and Hirumi, 2020). However, we also decided to divide the content into five shorter videos to facilitate participant attention and reduce the likelihood of losing focus during extended viewing periods. By presenting the material in manageable segments, participants were able to take shorter breaks as needed, which helped maintain focus to the video content. Additionally, breaking the content into multiple videos allowed for multiple measurements of participants' emotional and mental state.

There were four versions of the videos: five videos with a calm narrator, five with a calm narrator and SLS, five with an enthusiastic narrator, and five with an enthusiastic narrator and SLS, making it 20 different videos in total. They were made in Microsoft PowerPoint, the narrations were processed and edited with the Audacity® audio software (Audacity Team, 2021), and the subtitles were added with the Kapwing© online video editor (Kapwing, 2021). The type of videos is very simple and was made using basic and widely available tools to make sure that the findings of the study apply to a type of video that can be made by every educator and content maker that has access to a computer and the internet.

In general, the videos were created following existing guidelines for designing multimedia presentations for people with low prior knowledge on the topic. The learning material was mainly static representational pictures or graphics with minimal text (in black) on an off-white background. We did not use any decorative visuals or sounds and signals such as highlight or arrows helped learners to pay attention to important details and significant information. The narrated text used simple and casual language in shorter sentences.

The learning content was developed in collaboration with experts in wood science, building physics, building information modelling, service life modelling, entomology, and wood-based construction. The first video lasted for 3:13 minutes and introduced the participants to wood as a material and the concept of service life. It presented the benefits of wood as a building material, such as having good mechanical properties, its' earthquake and fire safety, and wood being a practical, sustainable, and human health friendly material. In addition, it introduced participants to functional, safety and

appearance limits of buildings. The second video was 5:49 minutes long and explained different wood degradation processes, such as weathering, rot, and insect activity, how to recognize degradation, what are its requirements, causes, and consequences, and introduced three types of degradation control measures that were explained more in detail in the following three videos: selecting the right materials, protective design measures, and regular maintenance. The third video (6:38 minutes long in the enthusiastic conditions and 6:44 minutes long in the calm conditions) revolved around material selection and introduced learners to factors that need to be included in the decision-making process when selecting the material for a specific project. Furthermore, participants learned about material natural resistance, durability classes (EN 350), the difference between softwood and hardwood and between heartwood and sapwood, and finally, listened to what is modified and engineered wood. The fourth learning video was titled Protection by design and was 3:49 minutes long. In it, participants learned about the importance of exposure to weather conditions and different use classes of wood (EN 335). With images of good and bad examples, they also learned about several important design principles, such as how to prevent water contact with wood and limit the time of contact with water in case it cannot be avoided. The last video in the series (4:30 minutes long in the enthusiastic conditions and 4:37 minutes long in the calm conditions) revolved around maintenance. Participants were introduced to different types of coatings, how they work, their benefits and problems, and examples.

For the narrations, a woman with a Standard American English accent was chosen so learners would not be distracted from the content when listening to a foreign accent. The narrator read the script (words) in two versions: one with an enthusiastic voice and the other with a calm, neutral voice. She was instructed to express emotions realistically and without exaggeration to make the narrations as lifelike as possible and thus replicate authentic online learning videos. To refine the recordings, multiple versions were made based on feedback from the candidate. For the enthusiastic clips, the narrator was advised to use an uplifting intonation and make regular changes in tone and pitch. Conversely, for the calm version, the narrator maintained a relatively constant tone and pitch, as per Collins (1978). The audio clips were edited by the candidate with the goal of making both the calm and enthusiastic versions of the videos (roughly) the same length.

The pitch of the narration in the videos was analysed using Praat 6.3.18, an open-source program for analysing phonetics (Boersma and Weenink, 2023). We extracted approximately 30-second segments (initial and ending sentences) from each of the ten videos, five with an enthusiastic and five with a calm narration (refer to Appendix 2 for

a table presenting segment details such as segments' length and the pitch's mean, median, standard deviation, minimum, and maximum values). A significant difference was observed in the average pitch ($t(50) = 13.140$, $p < .001$, mean difference = 46.859, 95% CI [39.696 –54.021], $d = 3.645$, 95% CI [2.494–4.777]) and pitch standard deviations ($t(50) = 8.552$, $p < .001$, mean difference = 19.485, 95% CI [14.909–24.062], $d = 2.372$, 95% CI [1.512–3.212]) between the enthusiastic and calm segments. Specifically, the audio segments from videos featuring an enthusiastic narrator exhibited a significantly higher pitch ($M_{enthusiastic} = 235.245$ Hz, $SD_{enthusiastic} = 13.231$ Hz; $M_{calm} = 188.387$ Hz, $SD_{calm} = 12.472$ Hz) and greater pitch variability ($M_{enthusiastic} = 61.931$ Hz, $SD_{enthusiastic} = 9.018$ Hz; $M_{calm} = 42.446$ Hz, $SD_{calm} = 7.325$ Hz).

To isolate the effect of vocal prosody on participants' learning from the video, one of the study's goals was to investigate the impact of emotional tone conveyed solely through vocal cues. Thus, the lecturer was not visually displayed, such as in a video or as an animated pedagogical agent, to avoid the potential influence of nonverbal communication through facial expressions and body language on the results. Instead, the audio was added to the learning slides to solely examine the effect of vocal prosody on learning.

The videos with subtitles had the subtitles embedded into the video, so the viewers could not accidentally turn them off.

3.3.2.4 Instruments

Most of the survey included adapted questionnaires that have been used and validated in previous international studies, together with questions examining knowledge that were developed specifically for the purpose of the study. Two versions of the survey were made: a Slovene and a Norwegian version. For the Slovene version, two researchers translated the materials from the source language into Slovene and reconciled any discrepancies. The Norwegian version was adapted by a Norwegian researcher who translated the materials from an English version of the survey with close collaboration with the candidate who checked that all items in Slovene and Norwegian version had the same meaning.

The reliability of each instrument will be evaluated using McDonald's ω . While Cronbach's α is more commonly used as a measure of internal consistency, it has rigid assumptions and often underestimates reliability when tau equivalence is violated (McNeish, 2017; Revelle and Zinbarg, 2009). In such cases, McDonald's ω is a more appropriate measure, while if tau equivalence is met, McDonald's ω yields the same

results as Cronbach's α (McNeish, 2017), which is why the former will be reported throughout the dissertation.

The survey measured the following variables: demographic variables (gender, age, study program, and year), pre-existing knowledge of the topic, subjective pre-existing knowledge and experience level, prior interest in the topic, subjective and objective English language ability, emotional state, mental effort, cognitive load, narrator perception, interest, and knowledge of the learning topic.

Knowledge (pre-test and post-test): Two separate tests were developed for the experiments in collaboration with experts in wood science – a pre-test and a post-test. The questions were different in the pre- and post-test to avoid priming the subjects to the type of content, as they could become especially attentive to the content from the initial questions when viewing the videos, and they could respond to those questions more accurately if they would be repeated after the video.

The pre-test's purpose was to measure pre-existing knowledge on the subject matter before watching the learning videos and involved eight multiple choice questions (Appendix 5) on the topic of wood as a material that was not covered in the videos. The questions have four possible answers and an "I do not know" option to avoid guessing. The questions and answers were presented in the language of the survey (Slovene or Norwegian). Before the analysis, one point was assigned to the correct answer and zero points to a wrong answer/"I don't know" option, so the maximum amount of points one could get was 8. Participants received no feedback on whether their answers were correct or not.

Difficulty indexes of the pre-test questions are displayed in Table 5. The item difficulty index is calculated by dividing the number of students who answered a question correctly by the total number of students who took the test, with a higher index indicating an easier question. The difficulty indexes of pre-test questions are low, especially for pre-test question 4, but this is to be expected as most of the respondents were from social sciences study programs and were expected to have low prior knowledge on the topic. However, the pre-test also had low reliability ($\omega = 0.457$; $\omega_{Slo} = 0.278$; $\omega_{Nor} = 0.402$ – reliability of the Norwegian pre-test does not include the 4th pre-test question as it has no variability).

Table 5: Item difficulty indexes of pre-test questions in Study 1

Question	IDI
PT1	0.39
PT2	0.25
PT3	0.25
PT4	0.05
PT5	0.15
PT6	0.15
PT7	0.13
PT8	0.16

The post-test is similar to the pre-test but included questions on the topics that were covered by the videos. There were 29 multiple-choice questions (see Appendix 6) with four possible answers but no “I don’t know” option so the participants had to choose an answer. The post-test did not include the option “I don’t know” to encourage participants to think more deeply about their answers, which allowed them to guess. However, to gain insight into whether participants were guessing, after every post-test question there was an additional question asking participants to rate the level of confidence in their answer as a percentage. To ensure all participants had the same idea of what percentages mean, the question was worded as follows: “How confident are you that this answer is correct (percentage)? Given that there are 4 possible answers, 25% is a complete guess.” In total, 19 questions measured retention and 10 measured transfer. Due to the presence of some technical and professional terms in the videos, all questions and answers were presented both in the language of the survey (Slovene or Norwegian) and in English, the language of the videos. Again, no feedback was given to the participants regarding their answers. The order of the questions was fixed and the same for all participants. We assigned one point to a correct answer and zero points to an incorrect answer prior to analysis, so the maximum number of points one could get in the retention test was 19, 10 for the transfer test, and 29 points in total.

A small pre-study with 13 participants (8 female and 5 male, $M_{age} = 32.38$, $SD_{age} = 11.95$) from Slovenia who did not participate in any later experiments was made to verify the adequacy of the questions. Participants with lower subjective knowledge on the topic of wood ($M = 2.54$, $SD = 0.78$, max 7) and good subjective understanding of English ($M = 5.38$, $SD = 1.45$, max 7) watched the videos, answered the 29 questions, and were asked to comment on the difficulty and clarity of each question. Difficulty indexes or the proportion of participants that answered correctly on each item are presented in Table 6.

For multiple choice tasks, the ideal item difficulty is approximately 0.60, but in a test, the range of difficulty indexes should generally fall between 0.15 and 0.90 (Bucik, 1997). Since all items had a difficulty index between 0.23 and 0.85 (with an average IDI of 0.55) and no comments were provided by the participants of the pre-study, it was deemed that the questions were appropriate to use in further experiments.

Participants of Study 1 had the option to answer the post-test two times: during the first data collection session and a week after the initial session. The order of the questions was the same as the first time. Table 6 showcases the item difficulty indexes also in the first and second part of Study 1, together with confidence or certainty levels of participants who answered the question correctly. Similarly to the results from the pre-study, the item difficulty indexes ranged from 0.22 to 0.71 in the main part and from 0.19 to 0.84 in the delayed part of the study, indicating that the knowledge test was appropriate. Both retention and transfer questions had varied difficulty indexes. Although the overall difficulty index for the knowledge test was lower than the optimal 0.60, with the main part at 0.46 and the delayed part at 0.42, this should not affect the results of our study. Our focus is to determine whether there is a difference between groups, rather than accurately assess the knowledge gained after watching the videos. Therefore, our main concern with the test is to avoid a ceiling or floor effect, where a considerable proportion of participants achieve the highest or lowest possible score due to questions being too easy or difficult, making the measure unable of discriminating between subjects at either end of the spectrum and thus having an adverse impact on the results (Salkind, 2010).

Table 6: Item difficulty indexes and confidence levels of correct responses on post-test questions in Study 1

Question	Type of knowledge	Pre-study (N = 13)	Study 1 – part 1 (N = 224)		Study 1 – part 2 (N = 94)	
		IDI	IDI	Confidence	IDI	Confidence
R1	Retention	0.77	0.66	71.37%	0.67	59.44%
R2	Retention	0.77	0.70	61.56%	0.84	44.78%
R3	Retention	0.23	0.23	46.19%	0.26	43.71%
R4	Retention	0.69	0.34	45.53%	0.24	33.91%
R5	Retention	0.54	0.44	55.23%	0.39	39.57%
R6	Retention	0.62	0.47	57.08%	0.38	43.75%
R7	Retention	0.62	0.45	56.72%	0.35	42.94%
R8	Retention	0.46	0.55	50.58%	0.63	37.31%

R9	Retention	0.46	0.42	62.73%	0.39	53.11%
R10	Retention	0.62	0.55	60.96%	0.38	48.00%
R11	Retention	0.62	0.42	51.87%	0.44	48.00%
R12	Retention	0.46	0.29	75.92%	0.32	61.37%
R13	Retention	0.69	0.64	65.74%	0.47	48.75%
R14	Retention	0.62	0.37	53.38%	0.46	46.33%
R15	Retention	0.62	0.42	49.78%	0.34	47.13%
R16	Retention	0.31	0.58	55.38%	0.38	43.42%
R17	Retention	0.69	0.71	56.90%	0.51	49.31%
R18	Retention	0.54	0.33	62.68%	0.23	50.14%
R19	Retention	0.38	0.26	58.49%	0.24	38.04%
T1	Transfer	0.77	0.61	69.07%	0.59	51.53%
T2	Transfer	0.46	0.62	53.61%	0.63	55.08%
T3	Transfer	0.31	0.28	50.16%	0.24	44.52%
T4	Transfer	0.69	0.51	62.73%	0.40	49.61%
T5	Transfer	0.38	0.35	50.63%	0.32	52.70%
T6	Transfer	0.85	0.58	68.00%	0.54	50.20%
T7	Transfer	0.31	0.33	54.14%	0.31	37.31%
T8	Transfer	0.54	0.62	59.33%	0.70	49.86%
T9	Transfer	0.46	0.41	71.47%	0.22	55.00%
T10	Transfer	0.38	0.22	43.02%	0.19	38.06%

Note. IDI – item difficulty index

The knowledge test had acceptable levels of internal consistency in the first phase when looking at the whole sample, but low reliability when looking at the Slovenian and Norwegian sample separately ($\omega = 0.704$; $\omega_{Slo} = 0.545$; $\omega_{Nor} = 0.670$) (McNeish, 2017). When testing the retention ($\omega_1 = 0.604$, $\omega_{Slo1} = 0.446$, $\omega_{Nor1} = 0.627$) and transfer tests ($\omega_1 = 0.472$, $\omega_{Slo1} = 0.301$, $\omega_{Nor1} = 0.414$) separately, reliability was low.

The same test administered a week after the learning session had reliability approaching acceptable levels ($\omega_2 = 0.640$). Separate reliability coefficients for the Slovenian and Norwegian sample were not calculated as only three Norwegian students responded to the delayed post-test. Delayed retention ($\omega_2 = 0.519$) and transfer ($\omega_2 = 0.307$) tests separately had low reliability.

Subjective pre-existing knowledge, experience, and interest: Before the pre-test, participants were asked to rate their knowledge of wood as a building material before the current survey, the amount of experience they have working with wood, and their level

of interest in the topic, all on a 7-point scale (1 – Very low/I have never worked with wood/I am not interested at all, 4 – Moderate/I rarely work with wood/Neither interested nor not interested, 7 – Very high/I work with wood very often/Very interested). The question about the interest in the topic of using wood as a building material was repeated in the delayed testing session.

English language: For the English language block, participants were required to answer three questions and complete a short English test. The first question asked participants to indicate the total number of languages they understand, including their native language. Participants were asked to provide a numerical response between 1 and 20. Next, participants were asked to rank their understanding of English among the languages they know. To assist in answering this question, an example was provided: “If you indicated above that you understand 4 languages and you think you understand English better than the other two foreign languages but less than your mother language, please indicate the number 2.” Again, participants provided a numerical response between 1 and 20 for this question. Finally, the third question asked participants to rate their ability to understand spoken English using a scale from 1 (very low) to 7 (very high).

As an objective measure of English proficiency, the *Lexical Test for Advanced Learners of English* or *LexTALE* (Lemhöfer and Broersma, 2012) was used. LexTALE ($\omega = 0.816$; $\omega_{Slo} = 0.745$; $\omega_{Nor} = 0.896$) is a standardized test and has been found to be a good predictor of vocabulary knowledge as well as a good indicator of general English proficiency (as measured by more thorough and extensive proficiency tests, such as the TOEIC and the Quick Placement Test). In the test, participants are asked to decide whether the presented word is an existing English word or not. It comprises of 60 trials and takes approximately 3.5 minutes to complete, making it a quick, simple, and reliable way of testing for English proficiency that is better than self-ratings. While the instructions were translated into Slovene/Norwegian, the (non-)words of the test remained the same.

Emotional outcomes: Three scales were used to verify the affective state of the participants, two of them being single-item scales measuring the two dimensions of the circumplex model of *core affect* – arousal/activation level and pleasure/valence (Russell, 1980; Russell et al., 1989). These single-item scales were chosen as their brevity is a great advantage when numerous assessments need to be conducted within a limited time frame such as in our case. Both scales were administered six times – first just before watching the first video and then after watching each video. Despite being brief, both measures have demonstrated their reliability and validity in previous studies (Killgore, 1998; Russell et al., 1989). These scales are similar, but different from the scales used to measure perception of the instructor voice in the pre-study, as those scales focused

on the instructor whereas the scales used in this experiment focus on the core affect dimensions of the participants.

In their original form, the items for pleasure and arousal were designed as a single-item affect grid in which respondents rate their current mood with a mark on a grid made up from columns defining the pleasure-displeasure score and rows defining the arousal-sleepiness score. In our experiment, we instead used two single-item scales with examples to help participants understand the questions better. The first (valence; $\omega = 0.904$; $\omega_{Slo} = 0.897$; $\omega_{Nor} = 0.933$) question asked: "How pleasantly do you feel at the moment? Examples of unpleasant feelings are nervousness, frustration, boredom, or sadness, while examples of pleasant feelings are enthusiasm, joy, contentment, or relaxation." The second question (arousal; $\omega = 0.901$; $\omega_{Slo} = 0.901$; $\omega_{Nor} = 0.905$) asked: "What is your level of activation at the moment, regardless of whether the feeling is pleasant or unpleasant? Examples of low activation are relaxation, boredom, contentment, or sadness, and examples of higher activation are alertness, enthusiasm, nervousness, or frustration." Participants answered both questions on a 9-point Likert-type scale (1 – Extremely unpleasant/low activation, 2 – Very unpleasant/low activation, 3 – Unpleasant/Low activation, 4 – Somewhat unpleasant/low activation, 5 – Somewhere in between, 6 – Somewhat pleasant/high activation, 7 – Pleasant/High activation, 8 – Very pleasant/high activation, 9 – Extremely pleasant/high activation). Both individual measurements and average scores will be utilized when analysing results.

Another scale used to measure the difference in affective states of participants before and after the learning session was the *Positive Activation, Negative Activation and Valence Short Scale* (PANAVA-KS; Schallberger, 2005), based on the model of two general activation systems of affect (Watson and Tellegen, 1985). This scale was used in several similar experiments (e.g., Beege and Schneider, 2023; Schneider et al., 2022). It consists of three dimensions: positive activation (four items; $\omega = 0.839$, $\omega_{Slo} = 0.837$, $\omega_{Nor} = 0.851$), negative activation (four items; $\omega = 0.836$, $\omega_{Slo} = 0.833$, $\omega_{Nor} = 0.865$), and valence (two items; $\omega = 0.699$, $\omega_{Slo} = 0.690$, $\omega_{Nor} = 0.769$). In it, participants were asked to rate how they are feeling at the moment on a 7-point bipolar Likert-type scale ranging from -3 to +3 (e.g., "satisfied – dissatisfied"; "full of energy – no energy", "stressed – relaxed"). Students used the PANAVA-KS two times during the experiment: just before watching the first video (baseline measure) and after watching the last video. In the analysis, we focused on the difference between the two measures (e.g., PA score after the learning session – baseline PA score) to control for the baseline measures.

Instructor perception: Four scales were used to verify how the instructor is perceived by the participants. The first three scales were the same scales that were used in the

pre-study. The first one was a four item scale adapted from similar studies (e.g., Lawson and Mayer, 2021; Lawson et al., 2021), in which participants rated the extent to which they thought the narrator expressed four emotions: enthusiasm, calmness, frustration, and boredom on a 7-point rating scale. Next were two one item scales, asking participants to rate how pleasant or unpleasant and how passive or active was the narrator, this time on a 7-point scale instead of a 9-point scale to keep the rating scale similar throughout the experiment.

The novelty from the pre-study is the *Agent Persona Inventory – Revised* (API-R; Schroeder et al., 2017, 2018), which is a modified version of the original Agent Persona Inventory (Ryu and Baylor, 2005) that has been commonly used in experiments interested in the user perception of pedagogical agents (e.g., Colliot and Jamet, 2018; Li et al., 2019; Liew et al., 2020; Mayer and DaPra, 2012). Similarly to the original version, the API-R contains 25 items divided into four subscales measuring how subjects perceive four characteristics of the agent (or in our case, the speaker): their credibility (five items; $\omega = 0.870$; $\omega_{Slo} = 0.863$; $\omega_{Nor} = 0.887$), ability to facilitate learning (ten items; $\omega = 0.897$; $\omega_{Slo} = 0.890$; $\omega_{Nor} = 0.936$), how human-like they are (five items; $\omega = 0.907$; $\omega_{Slo} = 0.897$; $\omega_{Nor} = 0.941$), and how engaging the agent was (five items; $\omega = 0.932$; $\omega_{Slo} = 0.930$; $\omega_{Nor} = 0.940$). All four subscales had high levels of reliability. The new version of the scale changed seven items to make them more consistent with the underlying constructs of the subscales (Schroeder et al., 2017). The original version had a 5-point Likert-type scale (1 – Strongly disagree to 5 – Strongly agree), but we changed it to a 7-point scale (1 – Strongly disagree, 2 – Disagree, 3 – Somewhat disagree, 4 – Somewhere in between, 5 – Somewhat agree, 6 – Agree, 7 – Strongly agree) to make it consistent with all other scales throughout the experiment.

Cognitive outcomes: Again, two measures were used to assess the cognitive outcomes of the participants. The first was one item used to measure *subjective mental effort* of students (Paas, 1992). This item is the most commonly used subjective method for measuring cognitive load in multimedia learning and instruction research as it is very easy to implement and can be used in a variety of learning contexts (Korbach et al., 2017, 2018). The item in question asked students to rate the amount of mental effort they invested in understanding the content from the video on a 9-point scale (1 – Very, very low mental effort, 2 – Very low mental effort, 3 – Low mental effort, 4 – Rather low mental effort, 5 – Neither low nor a high mental effort, 6 – Rather high mental effort, 7 – High mental effort, 8 – Very high mental effort, 9 – Very, very high mental effort). The question was repeated five times ($\omega = 0.916$; $\omega_{Slo} = 0.913$; $\omega_{Nor} = 0.935$), together with

the valence and activation level items after each video. In the analysis, all the measurements will be used separately and as an average.

The second scale measuring subjective cognitive load that was used in the experiment was the *Cognitive Load Questionnaire* developed by Klepsch et al. (2017). This self-report questionnaire can be easily adapted to various learning topics and contexts so it has been used in multiple empirical studies (Krieglstein et al., 2022). It is also particularly useful as it reliably differentiates between different types of cognitive load (Klepsch and Seufert, 2020), namely intrinsic (two items; $\omega = 0.646$; $\omega_{Slo} = 0.557$; $\omega_{Nor} = 0.853$; e.g., "Learning from the videos was very complex"), extraneous (three items; $\omega = 0.776$; $\omega_{Slo} = 0.778$; $\omega_{Nor} = 0.691$; "The design of the learning videos was very inconvenient for learning"), and germane cognitive load (two items; $\omega = 0.489$; $\omega_{Slo} = 0.495$; $\omega_{Nor} = 0.581$; "I made an effort, not only to understand several details, but to understand the overall context."). In its original form, the questionnaire has an additional item for measuring germane cognitive load. However, the authors noted that this item should be used if the study requires purposefully varying GCL in the given learning material (e.g., by presenting prompts), which our study does not, so the item has been omitted. Participants rated their degree of agreement with the statements on a 7-point Likert-type scale.

Situational interest: To assess the interest in the topic that was induced by the videos we used an instrument by Rotgans and Schmidt (2011). It comprises of six items, one of them being reversely valued ($\omega = 0.860$; $\omega_{Slo} = 0.864$; $\omega_{Nor} = 0.816$). Participants answered how much they agree with each statement on a 7-point scale.

Intrinsic motivation: An 8-item self-reporting questionnaire by Isen and Reeve (2005) was used to assess the participants' motivation to watch the videos for their inherent value, based on one's interest or enjoyment. The questionnaire is commonly used in multimedia learning research (e.g., Shangguan et al., 2020; Um et al., 2012) and has been shown as having good internal consistency ($\omega = 0.917$; $\omega_{Slo} = 0.929$; $\omega_{Nor} = 0.890$). The items were adapted to reflect the context of the study (e.g., "The videos stimulated my curiosity"). Participants answered how much they agree with the statements on a 7-point scale.

Video experience: Lastly, participants rated their level of agreement with five statements about their experience with watching the videos that have been used in recent research on the effect of emotional design on learning from multimedia materials (e.g., Lawson et al., 2021a, 2021c). The items give preliminary information on cognitive and affective outcomes and ask participants about their attention, effort, enjoyment,

perceived material difficulty, and if they would like more lessons similar to the ones they experienced. All items were analysed separately.

3.3.2.5 Data collection

Experimental data was collected throughout 2022 in several sessions. A convenience sampling method was used to invite students to participate, which means it was not a random selection. The students were invited to participate via email, presentations, or through their professor during or after a lecture. We made sure that students knew that participation was voluntary and that they could stop at any time without having to give a reason. Before the study began, participants read and agreed with an informed consent form. An ethical approval for the research from the Commission of the University of Primorska for Ethics in Human Subject Research was obtained prior to the beginning of the experiment. Students received no incentives for participation.

Each data collection session involved 5 to 20 students and lasted between 50 to 90 minutes. The whole procedure is represented in Figure 9. The survey was presented electronically on the online platform 1ka.si (Faculty of Social Sciences, University of Ljubljana, 2022). The design of the online survey required participants to complete all items on a particular scale before moving on to the subsequent section of the questionnaire to insure no missing data. The study sessions took place in either a computer room with faculty computers or a classroom where participants utilized their own laptops to watch videos and complete the survey. The researcher briefly introduced the study without revealing the independent variables and was present throughout the experiment to answer any questions, but other than that, participants performed the experiment individually and in their own pace.

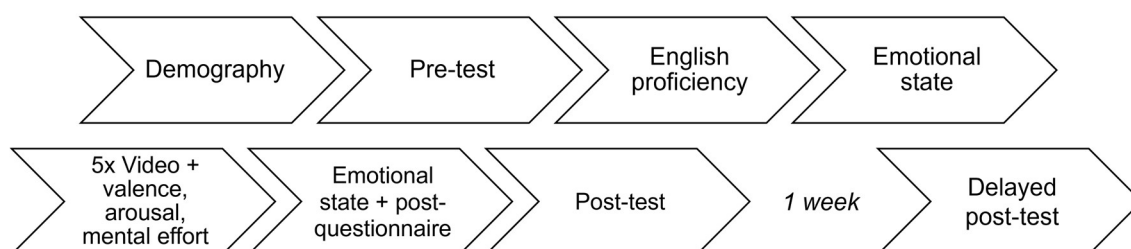


Figure 9: Experimental procedure of Study 1.

In every session, participants were randomly assigned to each experimental condition. The researcher prepared pieces of paper with links to each condition, counted the number of participants, and prepared the appropriate number of links in a bag.

Participants randomly drew the links to ensure that each group had the same number of participants in each condition.

All participants were provided with the same instructions for every data collection session, including detailed written instructions for each part of the study to ensure comparability of results from different sessions. For example, written instructions provided information on how to play the videos, such as the volume level, without rewinding, fast-forwarding, or pausing, at normal speed, with the same video quality, and with subtitles turned off. Before watching the experimental videos, there was also a short video test to ensure that participants could adjust the settings and identify potential problems that the candidate could solve before the independent variable was administered.

At the end of the survey, participants were asked to provide their student email and a 6-digit identification code. The identification code was created using the first two letters of their mother's name, the day of the month they were born, and the first two letters of their place of birth. This information was only used to send participants a link to the post-test a week after their initial participation and to connect their delayed post-test results with their initial results. Once the results were connected, the data with the email and identification code were deleted. Participants were sent only one email and were not contacted again. The email also included a thank you message for participating in the study, information on when the study results would be available, and an invitation to contact the candidate for more information.

3.3.2.6 Data analysis

Data was processed and analysed using Microsoft Excel, the open-source software jamovi (The jamovi project, 2022), and IBM SPSS Statistics 29.0.0.0 (IBM Corp., 2022). Descriptive statistics for all variables were computed, including a measure of central tendency (mean), dispersion (standard deviation), minimum and maximum answers, and the coefficients of kurtosis and skewness. Before conducting further analyses, boxplots were checked for outliers, and Shapiro-Wilk and Levene's tests were performed to check for assumptions of normality and homogeneity of variances (with results being displayed in Appendix 7). Then, groups were compared separately (2 by 2) based on the independent variable in focus by utilizing independent *t*-tests – parametric Student's *t*-tests when assumptions were met or only the normality assumption was violated and the non-parametric Welch's *t*-tests when all assumptions were not met. The central limit theorem suggests that when the sample size exceeds 30, the distribution tends to

approximate normality. Consequently, in the case of larger sample sizes, a violated normality assumption has minimal impact (Field, 2018), so when only the normality assumption was not met, the parametric test was still used. Mann-Whitney U test was used to discern differences in categorical demographic variables such as study program, study year, and country, and the chi-square (χ^2) test was utilised to test for differences in gender.

The experimental groups were then also compared using multivariate analysis of covariance (MANCOVA) and analysis of covariance (ANCOVA) to control for the effect of confounding variables. ANCOVA is a statistical technique that combines elements of analysis of variance (ANOVA) and regression, allowing us to assess group differences while controlling for the influence of one or more covariates, providing a more accurate assessment of the independent variable's impact (Navarro and Foxcroft, 2022). Before conducting ANCOVAs, homogeneity and normality tests were also performed to check whether assumptions for the test were met. In instances where ANCOVA assumptions were not satisfied, we used the Quade test, a non-parametric alternative to ANCOVA that uses a rank analysis (Quade, 1979). ANCOVA was also performed for comparing the groups on variables (emotional state) that included a baseline measure.

On the other hand, multivariate analysis of covariance or MANCOVA applies the same principles of ANCOVA but is used where there are multiple related outcomes. Huberty and Morris (1989) challenge the argument that performing a multivariate analysis of variance (MANOVA) as a precursor to multiple analyses of variance (ANOVAs) is needed to manage the risk of Type I error and argue that MANOVA and multiple ANOVAs address different research questions and that the results of one does not necessarily impact the results of the other. The use of MANOVA or MANCOVA is suggested only in cases where there is a good theoretical or basis for doing so (Field, 2018). However, due to the exploratory nature of our study, results of MANCOVA will also be reported in case of correlated dependent variables as it is able to detect smaller effects compared to ANCOVA and it can also examine the relationship between multiple dependent variables. The results of Box's homogeneity of covariance matrices tests and Shapiro-Wilk multivariate normality tests (assumption tests) will also be reported. A correlation matrix (Pearson r) is displayed in Appendix 3.

Lastly, multiple two-way ANOVAs were conducted to verify whether there is any interaction between the narrator emotional tone and SLS. Checking for interactions in statistical analyses is important because it helps explore whether there is a combined or joint effect of the independent variables on the outcome.

Due to a larger number of comparisons, a Bonferroni correction was used to mitigate the increased risk of Type I errors (Colman, 2014). However, marginally significant differences with a p -value of less than 0.10 will also be pointed out as a higher threshold can help in striking a balance between controlling Type I (false positives) and Type II (false negative) errors and is useful in exploratory research or when the expected effects are small to moderate, which is the case in this study. Due to the increased risk of chance results, interpretation of these kinds of findings will be made with caution.

Using the software G*Power (Faul et al., 2007), we calculated that with sample groups comprising 111 and 115 participants, an alpha level of 0.05, and a desired power of $1 - \beta = 0.80$, the study would have the capability to detect a moderate effect size (Cohen's d) of 0.37. This analysis assures us that the research design is sufficiently powered to identify meaningful differences between the two groups. In the delayed phase of the experiment, when the group sizes were reduced, namely to 47 and 47 for the narrator type and 48 and 46 for the SLS inclusion condition, the study had the statistical power to identify a large effect size of 0.58.

3.3.3 Results and interpretation

The chapter starts with an overview and comparison of the participant groups. Subsequently, the results and interpretation chapter is structured into sections based on the independent variables: narrator emotional tone and inclusion of same-language subtitles (SLS). Furthermore, within each section, there are subsections dedicated to the various types of dependent outcomes, namely instructor perception and emotional, cognitive, and learning outcomes. These subsections further incorporate the different variables measured during the experiment. In addition to presenting the actual analyses, each section also includes testing of assumptions (with numerical results being displayed in Appendix 7). The last section presents the results of multiple two-way ANOVAs and any potential interactions between narrator emotional tone and inclusion of SLS.

To account for multiple comparisons and reduce the likelihood of Type I errors, a Bonferroni correction was applied to all tests (Colman, 2014). For the group comparison prior to the intervention, an α level of 0.003 (0.05/17) was used, whereas for the comparison of groups on dependent variables, an α level of 0.002 (0.05/32) was utilized.

3.3.3.1 Groups' description and comparison

Prior to examining the findings of Study 1, we conducted preliminary analyses to assess potential group differences, as these variances might impact the alterations

observed in our dependent variables. First, there will be a description of the whole study sample, followed by descriptive statistics of control variables divided by groups and a comparison between groups. Instead of categorizing the sample into four groups (enthusiastic narrator without same-language subtitles, calm narrator without SLS, enthusiastic narrator with SLS, and calm narrator with SLS) and conducting multiple comparisons among them, we opted to present two comparisons based on factor (narrator's emotional tone and SLS inclusion) as it makes more sense for these data.

In general, participants rated their level of knowledge about wood as a building material at the start of the experiment as low to somewhat low ($M = 2.67$, $Mdn = 3$, $SD = 1.18$), which was supported by the results of the pre-test, where the average score was 1.54 ($Mdn = 1$, $SD = 1.37$) out of 8. Most participants did not have much experience working with wood ($M = 3.15$, $Mdn = 3$, $SD = 1.40$) and were neutral to learning about the subject – they felt they were neither interested nor not interested in the topic ($M = 3.98$, $Mdn = 4$, $SD = 1.58$). On average, students reported speaking 3 languages ($M = 3.20$, $Mdn = 3$, $SD = 1.88$) and self-rated their ability to understand spoken English as somewhat high ($M = 5.07$, $Mdn = 5$, $SD = 1.46$). They had an average score of 64.29 ($Mdn = 62.5$, $SD = 11.72$) out of 100 on the English vocabulary test.

Participants from each class were randomly assigned into each experimental group, thus ensuring an equal distribution (a matching number) of participants across the groups based on their country, study year, and study program. In the enthusiastic voice condition, there were 111 total participants, among them 92 women and 19 men, and 90 participants came from Slovenia and 21 from Norway, while in the calm voice condition, there were 115 participants – 91 women, 22 men, and two participants did not wish to disclose their gender, 94 were from Slovenia and 21 from Norway. A similarly equal distribution is present in the groups divided by the presence of SLS – there were 115 participants who saw videos without SLS, among them 86 women, 27 men, and two undisclosed; 94 were from Slovenia and 21 from Norway. On the other hand, 111 participants viewed videos with added SLS, 97 being women and 14 men, which is a smaller proportion compared to the group without SLS. 90 participants were from Slovenia and 21 from Norway.

Table 7 displays the descriptive statistics of the control variables and learners' characteristics separately for the groups with an enthusiastic voice and a calm voice, while Table 8 presents the same information, but divided by the conditions related to the presence of same-language subtitles.

Table 7: Learners' characteristics and descriptive statistics for variables before watching the videos divided by enthusiastic and calm conditions

	<i>M</i>		<i>SD</i>		Min–Max		Skewness		Kurtosis	
	Ent	Cal	Ent	Cal	Ent	Cal	Ent	Cal	Ent	Cal
SPK	2.65	2.70	1.13	1.23	1–6	1–6	0.54	0.20	0.20	–0.88
TPK	1.57	1.52	1.41	1.33	0–6	0–5	0.99	0.82	0.67	0.24
PE	3.08	3.21	1.33	1.44	1–6	1–6	0.46	0.32	–0.59	–0.91
PI	4.03	3.93	1.53	1.64	1–7	1–7	–0.05	–0.12	–1.01	–0.88
Lan	3.01	3.21	1.03	1.33	1–6	1–8	0.29	0.91	–0.06	1.60
SEP	5.11	5.03	1.49	1.44	1–7	2–7	–0.44	–0.41	–0.30	–0.66
TEP	64.10	64.47	11.41	12.05	42.5–91.3	37.5–100	0.61	0.65	–0.30	0.45
Val ^b	5.49	5.36	1.52	1.50	2–9	2–8	–0.08	0.04	–0.74	–0.83
AL ^b	4.66	4.71	1.51	1.61	1–9	1–8	0.04	–0.39	–0.34	–0.28
PA ^b	3.50	3.59	1.13	1.07	1–6.75	1.25–7	–0.01	0.32	–0.08	0.52
NA ^b	3.23	3.32	1.26	1.13	1–6.5	1–6.25	0.23	0.08	–0.54	–0.45
VA ^b	4.48	4.68	1.22	1.10	1.50–7	2–7	–0.23	–0.06	–0.09	–0.37
Age	20.25	20.53	2.01	3.15	18–34	18–45	3.77	5.05	21.30	33.40

Note. Ent – enthusiastic narrator group, Cal – calm narrator group; SPK – subjective prior knowledge, TPK – tested prior knowledge, PE – prior experience, PI – prior interest, Lan – number of spoken languages, SEP – subjective English proficiency, TEP – tested English proficiency, Val – valence, AL – activation level, ^b – baseline, PA – positive activation, NA – negative activation, VA – valence

Table 8: Learners' characteristics and descriptive statistics for variables before watching the videos divided by group without and with SLS

	<i>M</i>		<i>SD</i>		Min–Max		Skewness		Kurtosis	
	No SLS	SLS	No SLS	SLS	No SLS	SLS	No SLS	SLS	No SLS	SLS
SPK	2.70	2.65	1.19	1.18	1–6	1–6	0.36	0.35	–0.53	–0.33
TPK	1.43	1.67	1.19	1.52	0–5	0–6	0.86	0.84	0.44	0.13
PE	3.16	3.14	1.45	1.34	1–6	1–6	0.44	0.32	–0.70	–0.90
PI	4.15	3.80	1.52	1.64	1–7	1–7	–0.06	–0.08	–0.67	–1.19
Lan	3.10	3.13	1.05	1.33	1–6	1–8	0.41	0.95	–0.05	1.89
SEP	5.17	4.95	1.45	1.46	1–7	1–7	–0.52	–0.33	–0.36	–0.54
TEP	64.21	64.38	11.90	11.58	37.5–100	42.5–96.3	0.60	0.67	0.25	–0.09
Val ^b	5.34	5.50	1.50	1.52	2–9	2–8	0.06	–0.11	–0.67	–0.88
AL ^b	4.77	4.60	1.57	1.55	1–9	1–7	–0.14	–0.26	–0.11	–0.57
PA ^b	3.62	3.46	1.00	1.19	1.25–7	1–6.75	0.32	0.09	0.68	–0.19
NA ^b	3.22	3.33	1.16	1.26	1–5.75	1–6.5	0.03	0.24	–0.67	–0.41
VA ^b	4.62	4.54	1.06	1.26	1.5–7	1.5–7	–0.02	–0.25	0.03	–0.38
Age	20.38	20.41	2.43	2.88	18–34	19–45	3.47	6.15	15.50	49.00

Note. No SLS – group without same-language subtitles, SLS – group with same-language subtitles; SPK – subjective prior knowledge, TPK – tested prior knowledge, PE – prior experience, PI – prior interest, Lan – number of spoken languages, SEP – subjective English proficiency, TEP – tested English proficiency, Val – valence, AL – activation level, ^b – baseline, PA – positive activation, NA – negative activation, VA – valence

Assumptions of homogeneity of variances and normality were checked with Levene's and Shapiro-Wilk tests, and plots were checked for potential outliers. When comparing groups based on the inclusion of same-language subtitles (SLS) or the emotional tone of the narrator, we found that the groups exhibited equal variances in all variables, but the assumption of normality was violated for almost all variables in both

cases. However, this is not an issue in larger sample sizes (Field, 2018), so we proceeded with the *t*-tests.

Table 9: *t*-tests, normality, and homogeneity tests' results comparing the enthusiastic vs. calm narrator groups and the no SLS vs. SLS groups

	Enthusiastic vs. calm voice				No SLS vs. SLS			
	<i>t</i>	<i>p</i>	<i>W</i>	<i>F</i>	<i>t</i>	<i>p</i>	<i>W</i>	<i>F</i>
Subjective PKn	−0.30	0.766	0.92*	2.80	0.30	0.766	0.92*	0.01
Tested PKn	0.25	0.802	0.89*	0.28	−1.33	0.186	0.91*	7.55
Prior experience	−0.69	0.493	0.93*	1.58	0.12	0.909	0.92*	0.26
Prior interest	0.46	0.648	0.95*	0.07	1.65	0.101	0.96*	2.90
Languages	−12.60	0.209	0.94*	8.06	−0.19	0.848	0.91*	3.94
Subjective EngP	0.42	0.673	0.92*	0.18	1.13	0.260	0.93*	0.19
Tested EngP	0.07	0.945	0.97*	0.01	−0.11	0.911	0.97*	0.01
Val baseline	0.65	0.519	0.95*	0.06	−0.82	0.412	0.96*	0.18
AL baseline	−0.27	0.790	0.96*	0.04	0.78	0.437	0.97*	0.09
PA baseline	−0.61	0.543	0.99	0.77	1.06	0.288	0.99*	3.49
NA baseline	−0.59	0.555	0.99	2.15	−0.69	0.494	0.99	0.71
VA baseline	−13.04	0.193	0.98*	0.42	0.56	0.579	0.98	3.23
Age	−0.79	0.432	0.55*	2.41	−0.06	0.949	0.54*	0.04

Note. $df_1 = 1$, $df_2 = 224$; *W* – Shapiro-Wilk test result, *F* – Levene test result, * $p < 0.003$; PKn – prior knowledge, EngP – English proficiency, Val – valence, AL – activation level, PA – positive activation, NA – negative activation, VA – valence

Table 9 presents the outcomes of *t*-tests, following the same format as previously employed, for comparing the two groups distinguished by the emotional tone of the narrator and the presence of SLS. As can be seen, there were no significant differences between groups in any of the variables in the table, demonstrating that the groups were equal before introducing the independent variables with the learning videos. There were also no significant differences in study year (narrator emotional tone: $U = 6369$, $p = 0.974$; SLS inclusion: $U = 6371$, $p = 0.978$), study program (narrator emotional tone: $U = 6244$, $p = 0.774$; SLS inclusion: $U = 6352$, $p = 0.949$), country (narrator emotional tone: $U = 6341$, $p = 0.900$; SLS inclusion: $U = 6341$, $p = 0.900$), and gender (narrator emotional tone: $\chi^2(2, N = 226) = 2.15$, $p = 0.340$; SLS inclusion: $\chi^2(2, N = 226) = 6.71$, $p = 0.035$)

in both sets of conditions. Taken together, we can conclude that the groups were similar in basic characteristics and in potentially confounding variables before watching the videos.

We also checked for potential differences between the Slovenian and Norway sample. While they were not included in further analyses since both groups were equally represented in each of the experimental groups, it helps understand our sample better. The sample from Norway ($M = 23.00$, $SD = 4.46$) was significantly older on average ($t(224) = 7.98$, $p < .001$, mean difference = 3.20, 95% CI [2.41–3.99], $d = 1.36$, 95% CI [0.91–1.81]) than students from Slovenia ($M = 19.80$, $SD = 1.51$), and the Norwegian sample had a much more equal distribution regarding genders than the Slovenian sample ($\chi^2(2, N = 226) = 23.10$, $p < .001$), where most participants were women (86.96% compared to 54.76% from Norway). Participants from the Slovenian university were students from social science programs while participants from the Norwegian university came from various programs from the life sciences. While there was no significant difference in self-assessed prior knowledge ($t(224) = 1.56$, $p = 0.120$) and experience with the subject of wood as a building material ($t(224) = -0.26$, $p = 0.794$), there was a significant difference on the pre-test that demonstrated the actual prior knowledge ($t(224) = 10.57$, $p < .001$, mean difference = 2.02, 95% CI [1.65–2.40], $d = 1.81$, 95% CI [1.29–2.32]) and in interest in the subject ($t(224) = 7.60$, $p < .001$, mean difference = 1.84, 95% CI [1.36–2.32], $d = 1.30$, 95% CI [0.86–1.73]), with students from NMBU being more informed ($M = 3.19$, $SD = 1.23$) and interested in the topic ($M = 5.48$, $SD = 1.02$) than students from Slovenia ($M = 1.17$, $SD = 1.09$; $M = 3.64$, $SD = 1.49$). The Norwegian sample also had better subjective ($M = 6.24$, $SD = 1.06$; $t(224) = 6.24$, $p < .001$, mean difference = 1.44, 95% CI [0.99–1.89], $d = 1.07$, 95% CI [0.66–1.47]) and tested English proficiency ($M = 75.12$, $SD = 13.48$; $t(224) = 7.38$, $p < .001$, mean difference = 13.30, 95% CI [9.75–16.85], $d = 1.26$, 95% CI [0.83–1.69]) than the Slovenian sample ($M = 4.80$, $SD = 1.41$; $M = 61.82$, $SD = 9.75$). There were no differences in baseline emotional variables levels before watching the videos, like valence, activation, positive and negative activation.

When looking at differences between experimental groups divided by countries, a similar pattern emerges as when considering all the results together. Specifically, in both the Slovenian and Norwegian sample, there were no significant differences between groups with varying emotional tone or SLS inclusion in variables prior to watching the videos.

3.3.3.2 Emotional tone of the narrator

Instructor perception

Recognizing the emotion from the voice of the narrator

The first stage of the cognitive-affective model of e-learning involves the recognition of the instructor's emotions by the learners (Mayer, 2020), so we first asked participants to rate the narrator's emotions as a direct measure and the narrator's pleasantness and activation level as an indirect measure. First, we present the descriptive statistics of all the variables (Table 10), which are followed by assumptions' tests and the actual comparison between the groups who listened to an enthusiastic or calm narrator.

Table 10: Descriptive statistics for variables related to recognizing the narrator's emotion for enthusiastic and calm narrator groups

	<i>M</i>		<i>SD</i>		Min–Max		Skewness		Kurtosis	
	Ent	Cal	Ent	Cal	Ent	Cal	Ent	Cal	Ent	Cal
ENTH	3.54	2.12	1.62	1.48	1–7	1–7	–0.10	1.17	–0.81	0.50
CALM	5.35	6.01	1.30	1.27	2–7	1–7	–0.33	–1.41	–0.69	1.88
FRU	1.53	1.57	0.98	1.12	1–6	1–6	2.06	2.12	4.23	3.96
BOR	3.17	4.33	1.74	1.89	1–7	1–7	0.55	–0.14	–0.47	–0.89
PL	4.62	4.16	1.32	1.52	1–7	1–7	–0.38	–0.10	–0.01	–0.67
AL	3.78	3.02	1.38	1.48	1–7	1–7	–0.02	0.33	–0.37	–0.57

Note. Ent – enthusiastic narrator group, Cal – calm narrator group; ENTH – enthusiasm, CALM – calmness, FRU – frustration, BOR – boredom, PL – pleasantness, AL – activation level

Before comparing the two groups in terms of ratings, we first checked the assumptions for the independent and dependent *t*-tests by performing the Levene's homogeneity test, the Shapiro-Wilk normality test, and checking box plots for outliers. Results of the two tests are presented in Appendix 7.

Both groups had equal variances on all ratings, but all variables had also violated the assumption of normality. However, when sample size is bigger than 30, the central limit theorem posits that the distribution will approximate normality (Field, 2018). For this reason, the violated assumption of normality has little effect on large samples, and we will proceed with *t*-tests.

Table 11: Comparisons of the enthusiastic and calm narrator groups on discrete emotions, pleasantness, and activation level rating of the narrator using *t*-tests

	<i>t</i>	<i>p</i>	Mean difference	95% CI	<i>d</i>	95% CI
ENTH	6.894	< .001	1.419	1.10–1.82	0.917	0.63–1.20
CALM	–3.842	< .001	–0.657	–1.00—0.32	–0.511	–0.78—0.24
FRU	–0.241	0.810	–0.034	–0.31–0.24	–0.032	–0.29–0.23
BOR	–4.790	< .001	–1.159	–1.64—0.68	–0.637	–0.91—0.36
PL	2.452	0.015	0.465	0.09–0.84	0.326	0.06–0.59
AL	4.016	< .001	0.766	0.39–1.14	0.534	0.26–0.80

Note. *df* = 224, *CI* – confidence interval, *d* – effect size; ENTH – enthusiasm, CALM – calmness, FRU – frustration, BOR – boredom, PL – pleasantness, AL – activation level

As Table 11 shows, there was a moderate to large significant difference in how the group with the enthusiastic narrator and the group with the calm narrator rated the narrator's enthusiasm, calmness, boredom, and activation level, with the former group rating the narrator significantly higher in enthusiasm ($M_{ent} = 3.54$, $SD_{ent} = 1.62$; $M_{calm} = 2.12$, $SD_{calm} = 1.48$) and activation level ($M_{ent} = 3.78$, $SD_{ent} = 1.38$; $M_{calm} = 3.02$, $SD_{calm} = 1.48$), and the group with the calm narrator giving significantly higher ratings on the calmness ($M_{ent} = 5.35$, $SD_{ent} = 1.30$; $M_{calm} = 6.01$, $SD_{calm} = 1.27$) and boredom items ($M_{ent} = 1.53$, $SD_{ent} = 1.74$; $M_{calm} = 1.57$, $SD_{calm} = 1.89$), supporting Hypothesis 1, as well as the results from Pre-study 1. In the pre-study, the same group of participants compared and rated different videos, making it easier to spot the difference in tone between the enthusiastic and calm narrator. In this experiment, on the other hand, learners were exposed to only one version of the narrator – either enthusiastic or calm. The independent ratings indicate that our two types of interventions were distinct enough in their displayed arousal to potentially elicit different responses from the participants. Although there was also a slight difference in terms of the narrator's pleasantness between the two groups ($M_{ent} = 4.62$, $SD_{ent} = 1.32$; $M_{calm} = 4.16$, $SD_{calm} = 1.52$), the observed difference did not reach the corrected significance levels, suggesting that the two narrations did not significantly differ in terms of perceived valence. Overall, these results align with previous findings that learners are equally adept at recognizing emotional tone in voice alone as they are when an onscreen instructor offers additional social cues (Lawson and Mayer, 2021), even when the expressed emotion differed only in activation level and not in valence.

To control for the impact of confounding variables, such as prior interest in the topic, prior knowledge, English proficiency, and initial emotional state, a MANCOVA and additional ANCOVAs were performed. For the emotional state, the PANAVA-KS baseline measures were used instead of the single activation level and valence items. This choice was made because the three PANAVA-KS subscales offer a more comprehensive overview of participants' emotional states before viewing the videos, in contrast to the latter, which rely on just one item for each variable and thus provide less detailed information.

Before conducting singular ANCOVAs, a MANCOVA was conducted, together with Box's test and the Shapiro-Wilk test as assumption checks. The first showed that the assumption of homogeneity of covariance matrices is met ($\chi^2(21) = 26.616, p = 0.184$), while the second indicates a violation of the normality assumption ($W = 0.900, p < .001$). However, due to the large sample size, the violation of the assumption will be disregarded, and we will proceed with MANCOVA. The MANCOVA revealed a significant overall effect of the narrator emotion on the perceived narrator emotion, Wilks' Lambda = 0.723, $F(6, 213) = 13.595, p < .001$. As individual results of the MANCOVA are similar to the ones of singular ANCOVAs, only the latter will be reported.

Table 12 displays the results of multiple ANCOVAs, together with assumption checks and post-hoc test results. While the assumption of normality was not met for some of the variables, the assumption of homogeneity of variances was met for all, so the use of ANCOVAs was warranted. The inclusion of covariates did not change the results significantly, as a main effect and significant differences can be seen in the case of ratings of narrator's enthusiasm, calmness, boredom, activation level, and marginally pleasantness.

Table 12: ANCOVA and post-hoc comparisons of the enthusiastic and calm voice groups on discrete emotions, pleasantness and activation level rating of the narrator

	ANCOVA*			Post-hoc test		Assumption checks	
	<i>F</i>	<i>p</i>	η^2p	<i>t</i>	<i>Mean difference</i>	<i>F</i>	<i>W</i>
Enthusiasm	44.758	< .001	0.170	6.690	1.390	1.232	0.973*
Calmness	15.967	< .001	0.068	-3.996	-0.682	0.743	0.946*
Frustration	0.065	0.799	0.000	-0.255	-0.035	1.391	0.799*
Boredom	24.942	< .001	0.103	-4.994	-1.190	0.317	0.985*

Pleasantness	7.928	0.005	0.035	2.816	0.509	1.988	0.993
Activation level	17.297	< .001	0.074	4.159	0.764	0.236	0.993

Note. ANCOVA: $df_1 = 1$, $df_2 = 218$; Levene's test: $df_1 = 1$, $df_2 = 224$; assumption checks (Levene's test and Shapiro-Wilk test): * $p < .001$

Additionally, paired samples t -tests were also conducted to compare the enthusiasm and calmness ratings separately for the two groups. The group that watched the videos with the calm narrator gave the highest rating on the calmness item ($M_{calm} = 6.01$, $SD_{calm} = 1.27$), which was significantly higher ($t(114) = 20.80$, $p < .001$, mean difference = 3.89, 95% CI [3.52–4.26], $d = 1.94$, 95% CI [1.62–2.24]) than the rating on the enthusiasm item ($M_{calm} = 2.12$, $SD_{calm} = 1.48$). However, in the enthusiastic group, the highest rating was not for enthusiasm ($M_{ent} = 3.54$, $SD_{ent} = 1.62$), but for calmness ($M_{ent} = 5.35$, $SD_{ent} = 1.30$), with the difference also being significant ($t(110) = -9.89$, $p < .001$, mean difference = -1.81, 95% CI [-2.17—1.45], $d = -0.94$, 95% CI [-1.16—0.71]), contradicting the results that lead to the confirmation of the first hypothesis. This implies that although the participants who listened to the enthusiastic narrator perceived the narrator as more enthusiastic compared to the group that listened to the calm narrator, the enthusiastic narrator was actually perceived as more calm than enthusiastic, which may have an impact on the subsequent results.

Social partnership with the narrator

According to the second step in the cognitive affective model of e-learning, learners are expected to experience a greater sense of social connection with the instructor (Mayer, 2020). Previous research with onscreen pedagogical agents has shown that this feeling of social partnership is stronger with instructors displaying a more positive demeanor and with instructors who are more active (Lawson et al., 2021c), but we predicted this will also hold true when the emotion (and their activity level) is expressed only through voice (Hypothesis 2). There were four components of the Agent Persona Inventory – Revised (Schroeder et al., 2017, 2018) that were measured: the ability of the narrator to facilitate learning, their credibility, how human-like are they perceived, and their level of engagement. There was a moderate to high correlation between the four variables ($0.398 < r < 0.724$, $p < .001$) (Appendix 3). Descriptive statistics divided by group are reported in Table 13.

Table 13: Descriptive statistics for API components for enthusiastic and calm narrator groups

	<i>M</i>		<i>SD</i>		Min–Max		Skewness		Kurtosis	
	Ent	Cal	Ent	Cal	Ent	Cal	Ent	Cal	Ent	Cal
FL	4.23	3.79	0.91	1.22	1.7–6.2	1–6	–0.37	–0.10	0.13	–0.41
CR	5.28	4.80	0.80	1.21	3–7	1–7	–0.39	–0.76	0.05	0.77
HL	4.17	3.41	1.41	1.45	1–6.8	1–6.8	–0.45	0.11	–0.65	–0.75
EN	3.32	2.73	1.31	1.42	1–7	1–6.8	0.28	0.67	–0.34	–0.26

Note. Ent – enthusiastic narrator group, Cal – calm narrator group; FL – facilitating learning, CR – credibility, HL – human-likeness, EN – engaging

The assumption of equal variances was violated for the components “Facilitating learning” and “Credibility”, but not for “Human-like” and “Engaging”. Additionally, the assumption of normality was violated for the components “Credibility”, “Human-like” and “Engaging”, but not for “Facilitating learning”. No outliers were identified.

Due to the violated assumption of equal variances in two cases, the Welch’s *t*-test, a test that accounts for heterogeneous variances, was used instead of the Student’s *t*-test for all four variables. The Welch *t*-test was chosen instead of the Mann-Whitney U test because previous research suggests that an unequal variance *t*-test performs equally well as the Mann-Whitney U test in controlling Type I errors when variances are equal, and it performs even better than the U test when variances are unequal (Ruxton, 2006; Zimmerman and Zumbo, 1993). On the other hand, when the assumption of homogeneity of variance is violated, Welch’s *t*-test offers superior control over Type I error rates compared to the Student’s *t*-test and it maintains its robustness and performs similarly to Student’s *t*-test when the assumptions are satisfied (Delacre et al., 2017). Results of the tests are reported in Table 14.

Table 14: Comparison of the enthusiastic and calm voice groups on the Facilitating learning, Human-likeness, Credibility, and Engaging variables using Welch's *t*-tests

	<i>t</i>	<i>df</i>	<i>p</i>	<i>Mean difference</i>	<i>95% CI</i>	<i>d</i>	<i>95% CI</i>
Facilitating learning	3.113	210.50	0.002	0.443	0.16–0.73	0.413	0.15–0.68
Credibility	3.496	198.51	< .001	0.474	0.21–0.74	0.464	0.19–0.83
Human-likeness	4.040	223.99	< .001	0.768	0.39–1.14	0.537	0.27–0.81
Engaging	3.266	223.47	0.001	0.594	0.24–0.95	0.434	0.17–0.70

Note. *CI* – confidence interval, *d* – effect size

As can be deduced from Tables 13 and 14, participants who listened to the enthusiastic narrator rated the narrator significantly higher on all of API's components with a small to moderate effect size. They perceived that the enthusiastic narrator was more able to facilitate their learning, they were more credible, engaging, and human-like than the calm narrator. The results are therefore consistent with Hypothesis 2 predicting that the instructor displaying enthusiasm with their voice would be perceived as more credible, engaging, human-like, and as more able to help students learn, endorsing the second step of the cognitive affective model of e-learning. Previous research has shown that an onscreen instructor displaying active positive emotions is perceived more positively – as more credible, engaging, human-like, and being more able to facilitate students' learning – compared to an onscreen instructor conveying more passive positive emotions (Lawson et al., 2021a; Liew et al., 2020), but this study showed that the activation level of the expressed emotion can also lead to different perceptions of the instructor even if they are not visually present. However, this contrasts with results from a study on non-native speakers, which found no differences in instructor perception between weak and strong prosodic human voices (Davis et al., 2019). The inconsistency might be due to different voices used, different learning materials, or the different samples. While our study included participants with varying levels of English proficiency, the other study's sample consisted solely of Korean English (double) majors, whose proficiency levels are likely higher than the general non-native population. This suggests that even when focusing on non-native speakers, variations in these factors can lead to different outcomes.

The two groups were also compared on these variables including covariates. Both the Box's test ($\chi^2(10) = 32.864, p < .001$) and the Shapiro-Wilk test ($W = 0.959, p < .001$)

were significant, indicating a violation of assumptions of homogeneity of covariance matrices and multivariate normality. However, Box's test is notorious for both being susceptible to deviations from multivariate normality and for being significant in large samples, so it is common practice to disregard its' results when the compared sample sizes are equal (Field, 2018). The MANCOVA produced a significant effect of the narrator emotion on the variables measuring the participants' perception of the instructor (Wilks' Lambda = 0.902, $F(4, 215) = 5.816$, $p < .001$). Given the similarity between individual MANCOVA results and singular ANCOVAs, we will focus solely on the latter.

Table 15 presents the results of ANCOVAs and assumption tests for the four variables. As can be seen, two variables – “Facilitating learning” and “Credibility” – did not meet the homogeneity of variances assumption, so instead of ANCOVAs, ordinal logistic regression were performed. For the other two variables, “Human-like” and “Engaging,” the results of ANCOVAs are displayed. In both cases, a significant main effect was found, so post-hoc tests were performed, showing that even with the inclusion of covariates, the group that viewed the enthusiastic instructor considered them as significantly more human-like ($t(218) = 3.997$, $p < .001$, mean difference = 0.724, $d = 0.537$, 95% CI [0.267–0.806]) and engaging ($t(216) = 3.210$, $p = 0.002$, mean difference = 0.579, $d = 0.431$, 95% CI [0.163–0.699]) as their peers who watched the videos with the calm narrator.

Table 15: ANCOVA comparisons with five covariates of the enthusiastic and calm voice groups on the Facilitating learning, Human-likeness, Credibility, and Engaging variables

	ANCOVA*			Homogeneity test**		Normality test	
	<i>F</i>	<i>p</i>	η^2p	<i>F</i>	<i>p</i>	<i>W</i>	<i>p</i>
Facilitating learning				6.551	0.011	0.995	0.649
Credibility				15.027	< .001	0.973	< .001
Human-likeness	15.973	< .001	0.068	0.563	0.454	0.989	0.077
Engaging	10.305	0.002	0.045	0.574	0.449	0.977	0.001

Note. * $df_1 = 1$, $df_2 = 218$; ** $df_1 = 1$, $df_2 = 224$

As ANCOVA was not an option for the “Facilitating learning” and “Credibility” variables, the Quade non-parametric ANCOVA was used, showing that in both the “Facilitating learning” ($F(1,224) = 8.792$, $p = 0.003$, $t(224) = 2.065$) and “Credibility” ($F(1,224) = 9.406$, $p = 0.002$, $t(224) = 3.067$) variables there were significant differences between the two groups.

These results confirm that an enthusiastic narrator is perceived differently from a calm narrator, even after including covariates. This further supports the second step of the cognitive-affective model of e-learning when emotion is conveyed solely through voice.

Emotional outcomes

This section includes the following outcomes: differences in affective states as measured by PANAVA-KS and the valence and activation level questions, interest in the topic, intrinsic motivation to watch the videos, and learners' experience. As can be seen in Appendix 3, the different measures of affective state had a strong positive correlation ($0.526 < r < 0.685$, $p < .001$) and a low to high negative correlation with the negative activation scale ($-0.571 < r < -0.253$, $p < .001$).

Differences in affective state

Participants' affective states were assessed using three scales: the Positive Activation, Negative Activation and Valence Short Scale (PANAVA-KS), which participants completed before and after watching all the videos, and two single-item scales measuring participants' activation level and valence. The activation and valence scales were administered six times in total: once before watching the first video to establish a baseline, and after each of the five videos.

Since all the scales used in this study include a baseline measure, analyses of covariance (ANCOVA) were used to analyse the results. While both ANCOVA and the change from baseline method would be suitable in this context (where treatment assignment is randomized and independent of the baseline measurement), ANCOVA was selected due to its greater statistical power in randomized studies (Van Breukelen, 2006).

First, results from the PANAVA-KS will be presented (Table 16), followed by results on the activation level and valence scales.

Table 16: Descriptive statistics of PANAVA-KS values and change score for enthusiastic and calm narrator groups

	<i>M</i>		<i>SD</i>		Min–Max		Skewness		Kurtosis	
	Ent	Cal	Ent	Cal	Ent	Cal	Ent	Cal	Ent	Cal
PA ^b	3.50	3.59	1.13	1.07	1.00– 6.75	1.25– 7.00	– 0.01	0.32	– 0.08	0.52
PA ¹	2.91	3.11	1.14	1.22	1.00– 6.25	1.00– 7.00	0.32	0.38	– 0.29	0.13
PA ^c	–0.59	–0.47	1.25	0.95	–4.50– 2.25	–4.50 –2.00	– 0.38	– 0.80	0.70	3.02
NA ^b	3.23	3.32	1.26	1.13	1.00– 6.50	1.00– 6.25	0.23	0.08	– 0.54	– 0.45
NA ¹	3.07	3.02	1.18	0.92	1.00– 6.25	1.00– 5.00	– 0.02	– 0.27	– 0.43	– 0.57
NA ^c	–0.15	–0.30	0.93	0.99	–3.00– 2.25	–3.25 –2.00	– 0.34	– 0.62	0.50	0.43
VA ^b	4.48	4.68	1.22	1.10	1.50– 7.00	2.00– 7.00	– 0.23	– 0.06	– 0.09	– 0.37
VA ¹	4.24	4.44	1.11	1.02	1.00– 6.00	1.00– 7.00	– 0.39	– 0.03	– 0.10	0.40
VA ^c	–0.23	–0.24	1.16	1.13	–3.00– 2.50	–4.00 –3.00	– 0.08	– 0.49	0.09	1.59

Note. Ent – enthusiastic narrator group, Cal – calm narrator group; PA – positive activation, NA – negative activation, VA – valence, ^b – baseline, ¹ – measure after videos, ^c – change score (baseline measure subtracted from the measure after videos)

First, we compared differences in the two measures on the PANAVA-KS subscales, by performing paired samples *t*-tests separately in the group with the enthusiastic narrator and separately in the group with the calm narrator.

In the enthusiastic narrator group, the variables had a normal distribution. However, the same variables in the calm narrator group all violated the assumption of normality. Nevertheless, due to the sample size, we proceeded with parametric paired samples *t*-tests, the results of which are displayed in Table 17.

Table 17: Baseline and post-intervention differences in the PANAVA-KS measures for the enthusiastic narrator and the calm narrator groups separately

	<i>t</i>	<i>df</i>	<i>p</i>	<i>Mean difference</i>	<i>95% CI</i>	<i>d</i>	<i>95% CI</i>
Enthusiastic narrator group							
PA	4.940	110	< .001	0.586	0.35–0.82	0.469	0.27–0.66
NA	1.735	110	0.086	0.153	–0.02–0.33	0.165	–0.02–0.35
VA	2.121	110	0.036	0.234	0.02–0.45	0.201	0.01–0.39
Calm narrator group							
PA	5.349	114	< .001	0.474	0.30–0.65	0.499	0.30–0.69
NA	3.282	114	0.001	0.304	0.12–0.49	0.306	0.12–0.49
VA	2.227	114	0.028	0.235	0.03–0.44	0.208	0.02–0.39

Note. *CI* – confidence interval, *d* – effect size; PA – positive activation, NA – negative activation, VA – valence

In the enthusiastic narrator group, the only significant change from the baseline measure was in the positive activation. Specifically, the second measure of positive activation was significantly lower than their baseline, and the difference in the valence measure was approaching significance, with the post-intervention value also being lower than the baseline measurement. On the other hand, in the calm narrator group, both positive and negative activation post-intervention measures were significantly lower than the baseline, and the valence score after watching the videos with the calm narrator was also lower, but only approaching significance.

Before comparing these differences between the two narrator groups, assumption checks for ANCOVA were made by conducting Levene's and Shapiro-Wilk tests. No assumptions were violated for the positive activation scale ($W = 0.994$, $p = 0.436$; $F(1, 224) = 0.639$, $p = 0.425$) and the valence scale ($W = 0.989$, $p = 0.072$; $F(1, 224) = 0.149$, $p = 0.700$), but in the case of the negative activation scale, the distribution was not normal ($W = 0.983$, $p = 0.035$; $F(1, 224) = 0.657$, $p = 0.419$), which was not a problem due to the size of our sample.

Three ANCOVAs were made with the second measure of each PANAVA-KS subscale being a dependent variable and the baseline measure of each subscale being a covariate. Between the two groups, there was no significant difference in positive activation ($F(1, 223) = 1.261$, $p = 0.263$, $\eta^2 p = 0.006$), negative activation ($F(1, 223) =$

1.039, $p = 0.309$, $\eta^2p = 0.005$), and valence ($F(1, 223) = 0.814$, $p = 0.368$, $\eta^2p = 0.004$). No post-hoc tests were made due to these results.

To verify whether other confounding variables may affect the results, a MANCOVA and another three ANCOVAs were performed. The following variables were added as covariates: prior interest, tested prior knowledge, tested English proficiency, and all baseline measures from the PANAVA-KS instrument, meaning positive activation, negative activation, and valence. The MANCOVA did not yield a significant overall influence of narrator emotion on participants' emotional state, as indicated by Wilks' Lambda (0.984, $F(6, 213) = 1,169$, $p = 0.322$; $\chi^2(6) = 13.844$, $p = 0.031$, $W = 0.952$, $p < .001$).

Proceeding with singular tests, assumption checks revealed that ANCOVA is an appropriate test to use in all three cases (positive activation scale: $W = 0.992$, $p = 0.300$; $F(1, 224) = 0.770$, $p = 0.381$; negative activation scale: $W = 0.988$, $p = 0.047$; $F(1, 224) = 0.258$, $p = 0.612$; valence scale: $W = 0.987$, $p = 0.038$; $F(1, 224) = 1.263$, $p = 0.262$). Despite controlling for those variables, the results remained the same, as the difference between the two groups in positive activation ($F(1, 218) = 1.965$, $p = 0.162$, $\eta^2p = 0.009$), negative activation ($F(1, 218) = 0.730$, $p = 0.394$, $\eta^2p = 0.003$), and valence ($F(1, 218) = 1.536$, $p = 0.217$, $\eta^2p = 0.007$) remained insignificant. Again, no follow-up post-hoc tests were done.

While no differences were expected in the case of valence and negative activation, the non-existence of significant differences in the positive activation means that Hypothesis 3 cannot be substantiated. These findings suggest that the emotional tone of the narrator, whether enthusiastic or calm, does not significantly influence the participants' overall emotional state if the tone is conveyed only through voice. While previous studies have affirmed the impact of the emotional state of an instructor's emotional state on learners' emotions (Lawson et al., 2021a, 2021b, 2021c), our results indicate that the voice alone is not a sufficiently strong variable to affect these specific measures of emotional activation and valence.

In the following section, we will present the results of the activation level and valence single-item scales. These scales were administered both before and after watching each video. Descriptive statistics for these scales can be found in Table 18.

Table 18: Descriptive statistics of activation level and valence measurements and change score for enthusiastic and calm narrator groups

	<i>M</i>		<i>SD</i>		Min–Max		Skewness		Kurtosis	
	Ent	Cal	Ent	Cal	Ent	Cal	Ent	Cal	Ent	Cal
AL ^b	4.66	4.71	1.51	1.61	1–9	1–8	0.04	–0.40	–0.34	–0.28
AL ¹	4.77	4.90	1.50	1.55	1–9	1–8	–0.23	–0.40	–0.11	–0.17
AL ²	4.40	4.66	1.54	1.54	1–8	1–8	–0.27	–0.02	–0.53	–0.41
AL ³	4.09	4.27	1.60	1.66	1–7	1–8	–0.19	–0.09	–0.75	–0.55
AL ⁴	4.14	4.32	1.60	1.72	1–7	1–9	–0.24	–0.18	–0.60	–0.24
AL ⁵	4.14	4.29	1.73	1.79	1–9	1–9	0.01	–0.13	–0.31	–0.41
AL ^M	4.31	4.49	1.36	1.43	1–7	1–7.80	–0.32	–0.20	–0.55	–0.43
AL ^{cs}	–0.35	–0.22	1.49	1.45	4.40–4.20	4.40–4.00	0.07	0.01	0.94	1.05
V ^b	5.49	5.36	1.52	1.50	2–9	2–8	–0.08	0.04	–0.74	–0.83
V ¹	5.41	5.43	1.20	1.39	2–8	1–8	–0.31	–0.30	0.22	0.30
V ²	5.02	4.97	1.34	1.42	1–8	1–8	–0.45	–0.12	0.37	–0.28
V ³	4.69	4.67	1.46	1.50	1–8	1–8	–0.62	–0.22	0.23	–0.18
V ⁴	4.73	4.78	1.51	1.59	1–7	1–8	–0.47	–0.35	0.01	–0.05
V ⁵	4.65	4.73	1.63	1.67	1–9	1–9	–0.41	–0.41	0.20	0.22
V ^M	4.90	4.92	1.26	1.32	1.20–7.20	1.40–8.00	–0.52	–0.27	0.06	0.21
V ^{cs}	–0.59	–0.44	1.43	1.65	5.60–3.60	4.40–3.40	–0.34	–0.27	1.03	–0.01

Note. AL – activation level, V – valence, ^b – baseline, ^M – average of the five responses after watching each video, ^{cs} – change score (baseline measure subtracted from the average score)

As before, paired samples *t*-tests were performed separately in both groups to compare differences from baseline in the several measures. Most variables violated the

assumption of normality both in the case of the group with the enthusiastic narrator and in the group with the calm narrator.

Table 19: Baseline and post-intervention differences in the activation level and valence measures for the enthusiastic narrator and the calm narrator groups

	<i>t</i>	<i>p</i>	<i>Mean difference</i>	<i>95% CI</i>	<i>d</i>	<i>95% CI</i>
Enthusiastic narrator group*						
Activation I. ¹	-0.898	0.371	-0.117	-0.38–0.14	-0.085	-0.27–0.10
Activation I. ²	1.685	0.095	0.261	-0.05–0.57	0.160	-0.03–0.35
Activation I. ³	3.337	0.001	0.568	0.23–0.91	0.317	0.13–0.51
Activation I. ⁴	3.088	0.003	0.523	0.19–0.86	0.293	0.10–0.48
Activation I. ⁵	2.803	0.006	0.514	0.15–0.88	0.266	0.08–0.46
Activation I. ^M	2.470	0.015	0.350	0.07–0.63	0.234	0.05–0.42
Valence ¹	0.631	0.530	0.072	-0.15–0.30	0.060	-0.13–0.25
Valence ²	3.358	0.001	0.468	0.19–0.75	0.319	0.13–0.51
Valence ³	5.090	< .001	0.793	0.48–1.10	0.483	0.29–0.68
Valence ⁴	4.716	< .001	0.757	0.44–1.08	0.448	0.25–0.64
Valence ⁵	4.793	< .001	0.838	0.49–1.18	0.455	0.26–0.65
Valence ^M	4.310	< .001	0.586	0.32–0.86	0.409	0.22–0.60
Calm narrator group**						
Activation I. ¹	-1.329	0.187	-0.191	-0.48–0.09	-0.124	-0.31–0.06
Activation I. ²	0.374	0.709	0.052	-0.23–0.33	0.035	-0.15–0.22
Activation I. ³	2.959	0.004	0.443	0.15–0.74	0.276	0.09–0.46
Activation I. ⁴	2.231	0.028	0.391	0.04–0.74	0.208	0.02–0.39
Activation I. ⁵	2.522	0.013	0.426	0.09–0.76	0.235	0.05–0.42
Activation I. ^M	1.659	0.100	0.224	-0.04–0.49	0.155	-0.03–0.34
Valence ¹	-0.496	0.621	-0.070	-0.35–0.21	-0.046	-0.23–0.14
Valence ²	2.445	0.016	0.383	0.07–0.69	0.228	0.04–0.41
Valence ³	4.042	< .001	0.687	0.35–1.02	0.377	0.19–0.57
Valence ⁴	3.090	0.003	0.574	0.21–0.94	0.288	0.10–0.47
Valence ⁵	3.346	0.001	0.626	0.26–1.00	0.312	0.12–0.50

Valence ^M	2.866	0.005	0.440	0.14–0.74	0.267	0.08–0.45
----------------------	-------	-------	-------	-----------	-------	-----------

Note. * $df_1 = 110$, ** $df_2 = 114$, *CI* – confidence interval, *d* – effect size, ^M – average

Based on the findings presented in Table 19, it can be observed that in the enthusiastic narrator's group, the participants' activation level showed significant changes only after watching the third video, and even this change had a small effect size. However, videos 2 to 5, as well as the average post-intervention rating, led to significant changes in self-reported valence with small to medium effect sizes. In the calm narrator's group, the only significant change was also observed in valence after watching videos 3 to 5, albeit with a small effect size. These results indicate that watching the learning videos did not have a significant impact on participants' activation levels in either group. However, it is worth noting that both groups experienced a slightly more positive mood after learning from the videos compared to before watching them.

With ANCOVAs, we verified if there are potential differences between the groups in these ratings. Levene's and Shapiro-Wilk tests were also conducted to check for assumptions violations. In order to perform the ANCOVAs, activation level and valence baseline measurements were utilized as covariates, narrator emotion (enthusiastic vs. calm) was employed as the fixed factor, and the corresponding activation level/valence measurement (or average) was used as the dependent variable. Results of the twelve ANCOVAs and assumption checks are displayed in Table 20. As can be seen, contrary to Hypothesis 3, but similar to previous results with other measures of learners' emotional state, there were no significant differences between the groups in self-reported activation levels and valence after watching the learning videos, again showing that voice alone may not have the same effect on learners' emotions as an onscreen instructor. No post-hoc tests were made.

Table 20: ANCOVA comparisons of the enthusiastic and calm voice groups on activation level and valence items

	ANCOVA*			Homogeneity test**		Normality test	
	<i>F</i>	<i>p</i>	η^2p	<i>F</i>	<i>p</i>	<i>W</i>	<i>p</i>
Activation level ¹	0.345	0.557	0.002	0.163	0.387	0.980	0.002
Activation level ²	1.760	0.186	0.008	0.516	0.473	0.994	0.523
Activation level ³	0.621	0.432	0.003	0.450	0.503	0.987	0.041
Activation level ⁴	0.640	0.425	0.003	0.305	0.582	0.990	0.142
Activation level ⁵	0.301	0.584	0.001	0.229	0.633	0.993	0.370
Activation level ^M	0.946	0.332	0.004	0.100	0.753	0.988	0.066
Valence ¹	0.241	0.624	0.001	4.803	0.029	0.975	< .001
Valence ²	0.001	0.980	0.000	2.057	0.153	0.980	0.003
Valence ³	0.009	0.925	0.000	0.465	0.496	0.970	< .001
Valence ⁴	0.199	0.656	0.001	0.865	0.353	0.973	< .001
Valence ⁵	0.311	0.578	0.001	0.046	0.830	0.969	< .001
Valence ^M	0.146	0.703	0.001	0.655	0.419	0.973	< .001

Note. * $df_1 = 1$, $df_2 = 223$; ** $df_1 = 1$, $df_2 = 224$; ^M – average

A MANCOVA and a second set of ANCOVAs was made by adding the potentially confounding variables as covariates (prior interest, tested prior knowledge, tested English proficiency, and baseline measures of activation level and valence instead of the usual subscales from PANAVA-KS). The dependent variables in MANCOVA were all five activation level and valence measurements, but not the general measures. It did not reveal a significant effect (Wilks' Lambda = 0.979, $F(10, 210) = 0.454$, $p = 0.918$; $\chi^2(55) = 111.220$, $p < .001$, $W = 0.817$, $p < .001$).

Table 21 represents the results of multiple ANCOVAs and assumption checks (no serious violations were detected). Even with the incorporation of these additional variables, the findings remained unchanged, revealing no statistically significant differences. No further post-hoc tests were conducted.

Table 21: ANCOVA comparisons with five covariates of the enthusiastic and calm voice groups on activation level and valence items

	ANCOVA*			Homogeneity test**		Normality test	
	<i>F</i>	<i>p</i>	η^2p	<i>F</i>	<i>p</i>	<i>W</i>	<i>p</i>
Activation level ¹	0.517	0.473	0.002	0.788	0.376	0.982	0.006
Activation level ²	2.059	0.153	0.009	0.362	0.548	0.996	0.805
Activation level ³	0.776	0.379	0.004	0.460	0.498	0.988	0.062
Activation level ⁴	0.700	0.404	0.003	0.129	0.720	0.992	0.252
Activation level ⁵	0.347	0.556	0.002	0.482	0.488	0.995	0.678
Activation level ^M	1.145	0.286	0.005	0.027	0.870	0.988	0.059
Valence ¹	0.109	0.742	0.000	4.943	0.027	0.991	0.204
Valence ²	0.008	0.929	0.000	0.642	0.424	0.987	0.034
Valence ³	0.001	0.981	0.000	0.326	0.568	0.985	0.021
Valence ⁴	0.150	0.699	0.001	1.335	0.249	0.980	0.003
Valence ⁵	0.242	0.623	0.001	0.004	0.949	0.981	0.004
Valence ^M	0.073	0.788	0.000	0.583	0.446	0.977	0.001

Note. * $df_1 = 1$, $df_2 = 219$; ** $df_1 = 1$, $df_2 = 224$; ^M – average

Interest in the topic

We predicted that the enthusiastic narrator would trigger higher interest in the topic of using wood as a building material presented through the videos. Situational interest was measured with a short questionnaire after watching the videos and with the question “To what extent are you interested in the topic of using wood as a building material?” at the beginning of the delayed post-test the participants that 41.59% of participants took after a week. Both the questionnaire and the single item in the delayed part of the study had a non-normal distribution but equal variances, so a Student's *t*-test was used to compare results between the two groups.

There were no significant differences between participants listening to an enthusiastic an calm narrator in either situational interest ($t(223) = -0.494$, $p = 0.622$; $M_{enthusiastic} = 3.41$, $SD_{enthusiastic} = 1.15$; $M_{calm} = 3.49$, $SD_{calm} = 1.23$) and delayed interest ($t(92) = 0.409$, $p = 0.684$; $M_{enthusiastic} = 3.64$, $SD_{enthusiastic} = 1.36$; $M_{calm} = 3.53$, $SD_{calm} = 1.16$), failing to support Hypothesis 4 and challenging the cognitive affective model of e-learning when only the narrator's voice is used to convey emotion. An ANCOVA was

also performed for both measures to include prior interest, prior knowledge, English proficiency, and the three baseline measures of the PANAVA-KS as potentially confounding variables. While in the case of situational interest in the first experimental session the result did not change ($F(1, 217) = 0.623, p = 0.426, \eta^2p = 0.003; W = 0.995, p = 0.631; F(1, 223) = 0.235, p = 0.628$), a main effect was found in delayed interest ($F(1, 86) = 13.760, p < .001, \eta^2p = 0.138; W = 0.991, p = 0.766; F(1, 92) = 0.349, p = 0.556$), but a post-hoc comparison did not reveal any significant differences ($t(86) = 0.609, p = 0.544$, mean difference = 0.148, $d = 0.129$, 95% CI [-0.293–0.551]).

Intrinsic motivation

A more enthusiastic narrator could also help students be more motivated to learn about the topic from the videos, so motivation after the learning experience was compared between groups. Assumptions of normality and equal variances were both violated in this case, which is why a Welch's t -test was conducted. While no hypothesis was made in the case of intrinsic motivation due to lack of evidence, it was still predicted that those listening to an enthusiastic narrator would experience higher motivation than those listening to a calm and more neutral narrator. Contrary to the speculation, no significant differences were detected in level of motivation between the two groups ($t(223) = 1.222, p = 0.223; M_{enthusiastic} = 3.55, SD_{enthusiastic} = 1.13; M_{calm} = 3.35, SD_{calm} = 1.30$). As previously done, an ANCOVA was also made to control for the influence of prior interest, prior knowledge, English proficiency, and the baseline measures of positive activation, negative activation, and valence. In the case of the new model, both assumptions for ANCOVA were met ($W = 0.995, p = 0.636; F(1, 223) = 0.839, p = 0.361$). However, the results remained the same as before including the covariates, as there was no main effect ($F(1, 217) = 1.650, p = 0.200, \eta^2p = 0.008$).

Learners' experience

Based on the third step of the cognitive affective model of e-learning, after recognizing the narrator's emotions and feeling more social connection with them, learners should put more effort into learning the material (Mayer, 2020). A series of five questions commonly used in multimedia learning studies to measure learners' experience with the learning videos was employed, asking participants whether they were motivated to pay attention, how difficult the lectures were, how much effort did they exert to learn the information, how enjoyable was the experience, and if they would like more lessons like the one they just viewed. The correlation matrix in Appendix 3 shows

the relationships between the variables, with some not being significantly correlated (e.g., difficulty and paying attention, exerting more effort and enjoyment) while some being highly correlated (e.g., $r_{\text{enjoyment} - \text{paying attention}} = 0.690$, $p < .001$; $r_{\text{enjoyment} - \text{more lessons}} = 0.690$, $p < .001$). Descriptive statistics for these questions are displayed in Table 22.

Table 22: Descriptive statistics for the learners' experience questions for the enthusiastic and calm voice groups

	<i>M</i>		<i>SD</i>		Min–Max		Skewness		Kurtosis	
	Ent	Cal	Ent	Cal	Ent	Cal	Ent	Cal	Ent	Cal
PA	3.41	3.32	1.33	1.56	1–6	1–7	0.07	0.26	–0.69	–0.87
DIF	3.65	3.46	1.48	1.38	1–7	1–7	0.25	0.07	–0.53	–0.70
EF	3.59	3.56	1.40	1.46	1–7	1–6	–0.03	–0.05	–0.49	–1.01
ENJ	3.65	3.59	1.35	1.58	1–7	1–7	–0.11	–0.01	–0.49	–0.79
ML	3.41	3.19	1.44	1.66	1–7	1–7	–0.02	0.25	–0.54	–0.78

Note. Ent – enthusiastic narrator group, Cal – calm narrator group; PA – paying attention, DIF – difficulty, EF – exerting more effort, ENJ – enjoyment, ML – more lessons like this

Variables “Paying attention” and “Enjoyment” had unequal variances, while in the case of the variables “Difficulty”, “Exerting more effort”, and “More lessons”, the assumption of equal variances was not violated. All variables had a non-normal distribution. For simplicity's sake, Welch's *t*-test was performed for all five variables as in case of homogeneity of variances, the Welch's and Student's *t*-test's results are the same (Delacre et al., 2017).

Table 23: Comparison of the enthusiastic and calm voice groups on several variables on the learners' experience using Welch's *t*-tests

	<i>t</i>	<i>df</i>	<i>p</i>	<i>Mean difference</i>	<i>95% CI</i>	<i>d</i>	<i>95% CI</i>
PA	0.464	219.26	0.643	0.090	–0.29–0.47	0.062	–0.20–0.32
DIF	0.961	221.02	0.338	0.184	–0.19–0.56	0.128	–0.13–0.39
EF	0.126	223.95	0.899	0.024	–0.35–0.40	0.017	–0.25–0.28
ENJ	0.311	219.00	0.756	0.061	–0.33–0.45	0.041	–0.22–0.30
ML	1.025	220.15	0.306	0.212	–0.20–0.62	0.137	–0.13–0.40

Note. *CI* – confidence interval, *d* – effect size; PA – paying attention, DIF – difficulty, EF – exerting more effort, ENJ – enjoyment, ML – more lessons like this

As can be seen in Table 23, there were no significant differences between students who viewed videos with an enthusiastic and a calm narrator, failing to provide support the third step of the cognitive affective model of e-learning. These results indicate that, despite recognizing the narrator's emotions and feeling more social connection with the enthusiastic narrator, learners with the enthusiastic instructor did not have a significantly different learning experience compared to those who listened to the calm narrator.

Table 24: ANCOVA comparisons with six covariates of the enthusiastic and calm voice groups on several variables on the learners' experience

	ANCOVA*			Homogeneity test**		Normality test	
	<i>F</i>	<i>p</i>	η^2p	<i>F</i>	<i>p</i>	<i>W</i>	<i>p</i>
Paying attention	0.221	0.639	0.001	2.791	0.096	0.995	0.713
Difficulty	1.147	0.285	0.005	0.216	0.643	0.993	0.345
Exerting more effort	0.018	0.893	0.000	0.518	0.472	0.988	0.057
Enjoyment	0.051	0.821	0.000	0.005	0.942	0.997	0.901
More lessons like this	0.763	0.383	0.004	0.220	0.640	0.992	0.252

Note. * $df_1 = 1$, $df_2 = 217$; ** $df_1 = 1$, $df_2 = 223$

Additionally, a MANCOVA and five ANCOVAs were performed to include the effect of six confounding variables: prior interest, prior knowledge, English proficiency, and the three baseline measures of emotional state. The MANCOVA did not indicate a significant effect (Wilks' Lambda = 0.985, $F(5, 213) = 0.655$, $p = 0.658$; $\chi^2(15) = 15.020$, $p = 0.450$, $W = 0.954$, $p < .001$). Regarding ANCOVA, in the cases of all five variables, both ANCOVA assumptions were met (Table 24). Even after including covariates, the results remained the same, meaning that no main effect was observed, and no further analyses were made.

After consulting the correlation matrix in Appendix 3, the variables "Situational interest," "Intrinsic motivation," and "Enjoyment" showed a high correlation among them ($r_{\text{interest} - \text{motivation}} = r_{\text{interest} - \text{enjoyment}} = 0.811$, $r_{\text{motivation} - \text{enjoyment}} = 0.770$, $p < .001$), so a MANCOVA with these three outcomes was performed, producing a marginally significant effect (Wilks' Lambda = 0.961, $F(3, 215) = 2.912$, $p = 0.035$; $\chi^2(6) = 5.050$, $p = 0.537$, $W = 0.975$, $p < .001$). Conversely, as indicated by the results of ANCOVAs and *t*-tests, there was no significant effect of narrator emotion when looking at the "Situational interest"

($F(1, 217) = 0.352, p = 0.554$), “Intrinsic motivation” ($F(1, 217) = 2.218, p = 0.180$), and “Enjoyment” ($F(1, 217) = 0.131, p = 0.718$) variables separately.

Cognitive outcomes

Variables measuring cognitive outcomes include perceived cognitive load and mental effort. As can be observed from the correlational matrix in Appendix 3, the correlations between different types of cognitive load and the general mental effort measure ranged from insignificant to moderate ($-0.122 < r < 0.357$), solidifying the need to use different measures.

Cognitive load

The cognitive load questionnaire that was used differentiates between intrinsic, extraneous, and germane cognitive load, which is how the results will be presented. Based on the literature review, it was predicted that there will be a significant difference in (extraneous) cognitive load levels between the groups with a different narrator (Hypothesis 5). Table 25 presents descriptive statistics divided by group.

Table 25: Descriptive statistics of the cognitive load questionnaire for enthusiastic and calm narrator groups

	<i>M</i>		<i>SD</i>		Min–Max		Skewness		Kurtosis	
	Ent	Cal	Ent	Cal	Ent	Cal	Ent	Cal	Ent	Cal
ICL	4.05	3.96	1.20	1.29	2.00– 6.50	1.00– 7.00	0.14	–0.07	–0.70	–0.68
ECL	3.55	3.79	1.17	1.29	1.33– 6.67	1.00– 7.00	0.55	0.16	–0.05	–0.29
GCL	4.46	4.56	1.12	1.12	2.00– 6.50	1.00– 7.00	– 0.47	–0.55	–0.36	0.42

Note. Ent – enthusiastic narrator group, Cal – calm narrator group; ICL – intrinsic cognitive load, ECL – extraneous cognitive load, GCL – germane cognitive load

Based on the results of Levene's and Shapiro-Wilk tests, the assumption of homogeneity of variances was not violated in either case, while the assumption of normality was violated in all three cases.

The results of independent Student's *t*-tests fail to support Hypothesis 5, as there were no significant differences between the two groups in intrinsic ($t(223) = 0.54, p =$

0.593, mean difference = 0.09, 95% CI [-0.24–0.42], $d = 0.07$, 95% CI [-0.19–0.33]), extraneous ($t(223) = -1.46$, $p = 0.145$, mean difference = -0.24, 95% CI [-0.53–0.83], $d = -0.20$, 95% CI [-0.46–0.07]), and germane cognitive load ($t(223) = -0.68$, $p = 0.496$, mean difference = -0.10, 95% CI [-0.40–0.19], $d = -0.09$, 95% CI [-0.35–0.17]). These results mirror the findings of Davis et al. (2019), who observed higher germane cognitive load when comparing a weak prosodic human voice to a computer-generated voice, but found no differences when comparing non-native speakers learning from a strong and weak prosodic (human) voice.

The MANCOVA did not indicate a significant effect on the three cognitive load variables (Wilks' Lambda = 0.976, $F(3, 215) = 1.774$, $p = 0.153$; $\chi^2(6) = 1.822$, $p = 0.935$, $W = 0.981$, $p = 0.003$).

Additional ANCOVAs were made as the assumptions were met for all three variables ($W_{ICL} = 0.992$, $p_{ICL} = 0.229$, $F_{ICL}(1,223) = 1.055$, $p_{ICL} = 0.306$; $W_{ECL} = 0.989$, $p_{ECL} = 0.075$, $F_{ECL}(1,223) = 0.390$, $p_{ECL} = 0.533$; $W_{GCL} = 0.980$, $p_{GCL} = 0.003$, $F_{GCL}(1,223) = 0.022$, $p_{GCL} = 0.883$). The same results emerged even when accounting for prior knowledge, interest in the topic, English proficiency, and emotional state, as there was no significant result in the case of intrinsic ($F(1, 217) = 0.404$, $p = 0.526$, $\eta^2p = 0.002$), extraneous ($F(1, 217) = 2.020$, $p = 0.157$, $\eta^2p = 0.009$), and germane cognitive load ($F(1, 217) = 0.239$, $p = 0.625$, $\eta^2p = 0.001$).

Mental effort

After viewing each video, participants reported also the mental effort they invested into understanding the learning content, meaning that there are five separate mental effort measures. Descriptive statistics of the five measures, together with their average, are represented in Table 26.

Table 26: Descriptive statistics of the mental effort ratings for enthusiastic and calm narrator groups

	<i>M</i>		<i>SD</i>		Min–Max		Skewness		Kurtosis	
	Ent	Cal	Ent	Cal	Ent	Cal	Ent	Cal	Ent	Cal
ME ¹	4.76	4.59	1.70	1.59	1–8	1–9	-0.08	-0.01	-0.59	-0.26
ME ²	4.97	4.73	1.50	1.61	1–9	1–9	-0.28	-0.09	0.38	-0.35
ME ³	4.62	4.57	1.73	1.70	1–9	1–9	-0.05	-0.16	-0.31	-0.43

ME ⁴	4.54	4.51	1.63	1.70	1–9	1–9	–0.28	–0.36	0.15	–0.36
ME ⁵	4.58	4.50	1.70	1.75	1–9	1–9	–0.22	–0.33	0.31	–0.20
ME ^M	4.69	4.58	1.44	1.43	1.20– 8.60	1.20 – 8.80	–0.19	–0.16	0.25	0.04

Note. Ent – enthusiastic narrator group, Cal – calm narrator group; ME – mental effort, ^M – average

The assumption of homogeneity of variance was met for all variables, while the assumption of normality was violated for all variables.

Table 27: Comparison of the enthusiastic and calm voice groups on mental effort

	<i>t</i>	<i>p</i>	<i>Mean difference</i>	<i>95% CI</i>	<i>d</i>	<i>95% CI</i>
Mental effort ¹	0.756	0.451	0.166	–0.27–0.60	0.101	–0.16–0.36
Mental effort ²	1.170	0.243	0.243	–0.17–0.65	0.156	–0.11–0.42
Mental effort ³	0.247	0.805	0.056	–0.39–0.51	0.033	–0.23–0.29
Mental effort ⁴	0.124	0.901	0.028	–0.41–0.47	0.017	–0.24–0.28
Mental effort ⁵	0.314	0.754	0.072	–0.38–0.53	0.042	–0.22–0.30
Mental effort ^M	0.592	0.554	0.113	–0.26–0.49	0.079	–0.18–0.34

Note. *df* = 224, *CI* – confidence interval, *d* – effect size; ^M – average

Table 27 provides an overview of the results obtained from the comparison between the two groups regarding their mental effort levels. The table includes the data on all five mental effort items as well as their average scores. Upon analysing the data, it was found that there were no statistically significant differences observed between the group learning from an enthusiastic narrator and the group learning from a calm narrator in terms of their mental effort levels. This suggests that both groups exhibited similar levels of cognitive exertion during the learning process. The absence of significant differences implies that the choice of narrator style, whether enthusiastic or calm, did not significantly impact the participants' perceived mental effort. Similarly to the previous results, this finding again contradicts Hypothesis 5 regarding potential variations in mental effort levels based on narrator style.

A MANCOVA accounting for the participants' initial emotional state, English proficiency, and prior knowledge and interest in the topic and including all five individual

measurements of mental effort did not produce a significant effect (Wilks' Lambda = 0.989, $F(5, 214) = 0.482$, $p = 0.790$; $\chi^2(15) = 14.739$, $p = 0.470$, $W = 0.892$, $p < .001$).

While the normality assumption was not met for most of the variables, ANCOVAs were still conducted due to the large sample size. As can be deducted from Table 28, when accounting for the covariates, the results remained the same and no significant effect was found.

Table 28: ANCOVA comparisons with six covariates of the enthusiastic and calm voice groups on mental effort

	ANCOVA*			Homogeneity test**		Normality test	
	<i>F</i>	<i>p</i>	η^2p	<i>F</i>	<i>p</i>	<i>W</i>	<i>p</i>
Mental effort ¹	0.630	0.428	0.003	0.202	0.653	0.993	0.408
Mental effort ²	2.160	0.143	0.008	1.773	0.184	0.992	0.286
Mental effort ³	0.083	0.774	0.000	0.020	0.888	0.985	0.019
Mental effort ⁴	0.032	0.858	0.000	0.231	0.631	0.975	< .001
Mental effort ⁵	0.150	0.699	0.001	0.045	0.832	0.980	0.002
Mental effort ^M	0.499	0.481	0.002	0.000	0.989	0.981	0.004

Note. * $df_1 = 1$, $df_2 = 218$; ** $df_1 = 1$, $df_2 = 224$; ^M – average

Learning outcomes

In this section of the results, we examine a range of variables related to learning outcomes. These variables include tested knowledge, both in terms of retention and transfer, as well as participants' certainty regarding the correctness of their answers. Additionally, we explore participants' self-evaluation of their overall test performance. Furthermore, we extend our investigation to the delayed part of the experiment, where we assess delayed tested knowledge, certainty, and self-evaluation, offering valuable perspectives on the durability of acquired knowledge and any shifts in participants' self-evaluation and certainty over time. We will first present the results obtained from the main part of the experiment, shedding light on participants' learning progress during the immediate phase. Subsequently, we will delve into the outcomes stemming from the delayed part of the experiment, as well as report potential differences between the two testing sessions. At the end, results from the main and delayed parts of the experiments will also be compared.

Objective and subjective test performance in the immediate part of the experiment

Table 29 contains the descriptive statistics for all learning-related outcomes from the initial phase of the experiment. The variable "knowledge" includes the cumulative points earned on the test, while "retention" and "transfer" variables include points accumulated when correctly answering questions related to retention and transfer, respectively. These statistics provide a comprehensive snapshot of the participants' performance. There was a high correlation between retention and transfer ($r = 0.517$, $p < .001$) and a low to moderate correlation between self-evaluated test performance and the actual score on the retention ($r = 0.300$, $p < .001$) and transfer ($r = 0.287$, $p < .001$) part of the test (Appendix 3).

Table 29: Descriptive statistics of the learning outcomes from the immediate part of the experiment for enthusiastic and calm narrator groups

	<i>M</i>		<i>SD</i>		Min–Max		Skewness		Kurtosis	
	Ent	Cal	Ent	Cal	Ent	Cal	Ent	Cal	Ent	Cal
K	13.21	13.70	4.21	4.78	5–23	5–26	0.33	0.79	–0.47	0.11
R	8.82	8.97	2.93	3.37	2–16	3–17	0.26	0.64	–0.25	–0.26
T	4.39	4.73	1.91	2.03	0–8	0–10	0.02	0.24	–0.33	–0.18
C	52.43	51.63	17.23	22.65	8.62– 87.03	0.34– 95.17	–0.28	–0.23	–0.35	–0.66
C ^y	55.00	55.32	18.76	23.83	10.00 – 92.50	0.00– 95.91	–0.17	–0.34	–0.69	–0.74
C ⁿ	49.35	47.53	16.24	21.68	5.57– 79.21	0.00– 93.25	–0.49	–0.06	0.11	–0.45
RC	51.84	51.84	17.84	22.61	8.21– 90.00	0.53– 97.37	–0.18	–0.23	–0.50	–0.69
RC ^y	54.97	55.56	19.98	23.71	12.50 – 96.43	0.00– 100.0 0	–0.07	–0.31	–0.85	–0.74
RC ⁿ	48.33	46.96	16.84	21.37	2.56– 85.57	0.00– 91.00	–0.28	–0.10	–0.16	–0.51
TC	53.54	51.23	18.34	24.52	7.50– 91.50	0.00– 92.60	–0.32	–0.18	–0.41	–0.77

TC ^y	55.67	53.93	21.13	26.37	6.00– 100.0 0	0.00– 100.0 0	–0.06	–0.24	–0.73	–0.93
TC ⁿ	51.64	48.74	18.41	25.07	0.00– 100.0 0	0.00– 100.0 0	–0.24	0.13	0.01	–0.58
SE	3.20	3.46	1.10	1.23	1–6	1–7	–0.28	–0.06	0.24	0.61

Note. Ent – enthusiastic narrator group, Cal – calm narrator group; K – knowledge, R – retention, T – transfer, C – certainty, ^y – correct answers, ⁿ – incorrect answers, SE – self-evaluation

The results of Shapiro-Wilk's tests of normality and Levene's tests of homogeneity of variances (Appendix 7) reveal that some variables not only displayed violations of the assumption of normality but also exhibited discrepancies in homogeneity of variances between the groups. Due to these complexities, Welch's *t*-tests were used to compare the groups with enthusiastic and calm narrators, as this method accounts for unequal variances. In the interest of clarity and ease of interpretation, we applied Welch's *t*-tests on all learning variables.

Table 30: Comparison between the enthusiastic and calm voice groups on various learning variables in the immediate part of the experiment using Welch's *t*-tests

	<i>t</i>	<i>df</i>	<i>p</i>	<i>Mean difference</i>	<i>95% CI</i>	<i>d</i>	<i>95% CI</i>
K	–0.818	219.37	0.414	–0.492	–1.68–0.69	–0.109	–0.37–0.15
R	–0.364	218.77	0.716	–0.154	–0.99–0.68	–0.049	–0.31–0.21
T	–1.286	221.59	0.200	–0.338	–0.86–0.18	–0.172	–0.43–0.09
C	0.298	208.98	0.766	0.800	–4.49–6.09	0.040	–0.22–0.30
C ^y	–0.114	211.98	0.910	–0.325	–5.97–5.32	–0.015	–0.28–0.25
C ⁿ	0.712	207.54	0.477	1.820	–3.22–6.86	0.095	–0.17–0.36
RC	0.002	212.19	0.999	0.004	–5.36–5.36	0.000	–0.26–0.26
RC ^y	–0.202	217.01	0.840	–0.591	–6.36–5.18	–0.027	–0.29–0.24
RC ⁿ	0.535	212.07	0.593	1.373	–3.69–6.44	0.071	–0.19–0.33
TC	0.800	207.37	0.425	2.311	–3.39–8.01	0.107	–0.16–0.37
TC ^y	0.545	212.92	0.587	1.744	–4.57–8.06	0.073	–0.19–0.34
TC ⁿ	0.984	203.78	0.326	2.896	–2.91–8.70	0.132	–0.13–0.39

SE	-1.683	220.32	0.094	-0.262	-0.57-0.05	-0.225	-0.49-0.04
----	--------	--------	-------	--------	------------	--------	------------

Note. CI – confidence interval, *d* – effect size; K – knowledge, R – retention, T – transfer, C – certainty, ^y – correct answers, ⁿ – incorrect answers, SE – self-evaluation

The outcomes presented in Table 30 indicate that no statistically significant differences were observed between the two groups across any of the learning outcomes. This lack of significance extended even to variables measuring participants' certainty in their answers. These findings challenge Hypothesis 6, which had posited that there would be significant differences in knowledge levels between the group exposed to an enthusiastic narrator and the group exposed to a calm narrator. Despite the fact that previous studies affirmed the role of the instructor's emotions on learning outcomes (Lawson et al., 2021a, 2021b, 2021c), a disembodied instructor's voice may not have the same effect on learning from video. A study on non-native speakers reported similar findings (Davis et al., 2019), highlighting the importance of measuring various learning outcomes, as results may vary across different metrics. While they only measured retention, our study assessed multiple learning outcomes and did not find significant differences.

However, there was a marginally significant difference with a *p*-value of less than 0.1 – participants' subjective assessment of displayed knowledge during the test. While this result should be considered with caution as the chance of a Type II error is higher, it may also indicate that participants who viewed the videos with a calm narrator were slightly more optimistic regarding their test performance.

As was done in previous cases, additional tests were performed to look at the results when controlling for six confounding variables. First, a MANCOVA was conducted including some of the dependable variables, as a lot of the dependable variables in this section are computed. We included the separate results of the retention and transfer parts of the test, the level of certainty in their correct and incorrect answers, also divided by the retention and transfer parts of the test, and self-evaluated test performance, resulting in seven dependent variables. No significant effect was observed (Wilks' Lambda = 0.962, $F(7, 207) = 1.172$, $p = 0.320$; $\chi^2(28) = 45.928$, $p = 0.018$, $W = 0.884$, $p < .001$).

Following up with univariate tests, ANCOVA assumption checks were made for all variables and are displayed in Table 31, from which we can observe that most of the variables pertaining to certainty in one's answers did not meet the assumptions. For this reason, ANCOVAs were performed for the variables knowledge, retention, transfer, and

self-evaluation of test performance, while the variables connected to participants' certainty were analysed with the Quade non-parametric test.

Table 31: ANCOVA comparisons with six covariates of the enthusiastic and calm voice groups on various learning variables in the immediate part of the experiment

	ANCOVA*			Homogeneity test**		Normality test	
	<i>F</i>	<i>p</i>	η^2p	<i>F</i>	<i>p</i>	<i>W</i>	<i>p</i>
Knowledge	1.708	0.193	0.008	0.334	0.564	0.992	0.223
Retention	0.470	0.494	0.002	0.021	0.884	0.993	0.380
Transfer	2.971	0.086	0.014	1.038	0.309	0.995	0.632
Certainty				6.498	0.011	0.986	0.024
Certainty in correct answers				6.361	0.012	0.984	0.013
Certainty in incorrect answers				8.542	0.004	0.990	0.119
R Certainty				5.956	0.015	0.989	0.076
R Certainty in correct answers				2.433	0.120	0.986	0.024
R Certainty in incorrect answers				5.480	0.020	0.991	0.178
T Certainty				7.212	0.008	0.985	0.022
T Certainty in correct answers				6.912	0.009	0.992	0.256
T Certainty in incorrect answers				6.654	0.011	0.995	0.693
Self-evaluation	3.414	0.066	0.016	0.000	0.997	0.980	0.002

Note. * $df_1 = 1$, $df_2 = 216$; ** $df_1 = 1$, $df_2 = 222$; R – retention, T – transfer

ANCOVA results (Table 31) are similar to the results of the *t*-tests. However, two variables show marginally significant results ($p < 0.10$) – transfer and self-evaluation of one's test performance. Similarly as in the case above, for these two variables, post-hoc comparisons were performed. Participants who watched the videos with a calm narrator had marginally higher results on the transfer part of the knowledge test ($t(216) = -1.724$, $p = 0.086$, mean difference = -0.393 , $d = -0.232$, 95% *CI* $[-0.499-0.034]$) and reported better test performance ($t(216) = -1.848$, $p = 0.066$, mean difference = -0.273 , $d = -0.249$, 95% *CI* $[-0.516-0.018]$) than those who learned from videos with an enthusiastic

narrator. It has been shown that an emotional design intervention can affect transfer but not retention, but only when learners are aware of the intervention (Bender et al., 2021), which they were not in our study. Our findings, however, are just tentative and should be considered with caution.

As ANCOVA could not be performed for the several certainty level variables, the non-parametric ANCOVA was conducted. No variable showed a significant effect of the narrator emotion (Table 32).

Table 32: Quade test comparisons with six covariates of the enthusiastic and calm voice groups on various certainty level variables in the immediate part of the experiment

	<i>F</i>	<i>p</i>
Certainty	0.000	0.987
Certainty in correct answers	0.294	0.588
Certainty in incorrect answers	0.581	0.447
R Certainty	0.093	0.761
R Certainty in correct answers	0.298	0.586
R Certainty in incorrect answers	0.278	0.599
T Certainty	0.243	0.623
T Certainty in correct answers	0.018	0.892
T Certainty in incorrect answers	1.154	0.284

Note. $df_1 = 1$, $df_2 = 222$; R – retention, T – transfer

Objective and subjective test performance in the delayed part of the experiment

One week after the initial part of the experiment, a total of 94 participants solved the knowledge test once again. Of those, 47 participants watched videos with an enthusiastic instructor and 47 participants watched videos with a calm instructor. During this second assessment, they were asked to reevaluate their confidence in each answer and provide a subjective perception of their overall performance on the second test. The ensuing results will be presented in a manner consistent with the previous section, showcasing descriptive statistics in Table 33 for clarity and comparison.

Similar to the previous section, there was a high correlation between the retention and transfer part of the delayed test ($r = 0.568$, $p < .001$) and a low correlation between the scores on each part of the test and self-evaluated test performance ($r_{\text{self-evaluation} - \text{retention}} = 0.256$, $p = 0.013$; $r_{\text{self-evaluation} - \text{transfer}} = 0.305$, $p = 0.003$).

Table 33: Descriptive statistics of the learning outcomes from the delayed part of the experiment for enthusiastic and calm narrator groups ($N = 94$)

	<i>M</i>		<i>SD</i>		Min–Max		Skewness		Kurtosis	
	Ent	Calm	Ent	Calm	Ent	Calm	Ent	Calm	Ent	Calm
K	12.53	11.64	3.65	4.35	4–21	4–27	0.16	0.84	0.40	2.09
R	8.26	7.62	2.68	2.92	2–15	3–18	0.25	0.88	0.47	1.98
T	4.28	4.02	1.60	1.79	0–7	1–9	–0.28	0.49	–0.19	0.21
C	43.39	42.66	20.03	17.19	0.00–77.93	0.00–79.31	–0.27	–0.09	–0.57	–0.02
C ^y	45.76	44.54	20.61	17.98	0.00–82.56	0.00–82.56	–0.26	–0.13	–0.48	–0.24
C ⁿ	41.58	41.15	19.79	17.52	0.00–77.50	0.00–80.36	–0.22	0.02	–0.60	–0.02
RC	41.24	40.76	19.76	17.02	0.00–74.74	0.00–76.32	–0.16	–0.07	–0.64	–0.17
RC ^y	43.80	43.12	20.48	17.98	0.00–83.67	0.00–83.67	–0.12	–0.03	–0.58	–0.17
RC ⁿ	38.71	38.77	19.50	17.24	0.00–72.22	0.00–77.78	–0.11	0.12	–0.70	–0.17
TC	47.47	46.27	21.99	18.58	0.00–84.00	0.00–85.00	–0.25	–0.09	–0.63	–0.12
TC ^y	48.71	46.38	23.56	20.69	0.00–93.75	0.00–85.00	–0.17	0.15	–0.62	–0.62
TC ⁿ	46.34	44.95	22.21	19.49	0.00–90.00	0.00–85.00	–0.7	0.02	–0.57	–0.27
SE	3.02	3.09	1.09	1.14	1–5	1–5	–0.56	–0.27	–0.75	–1.04

Note. Ent – enthusiastic narrator group, Cal – calm narrator group; K – knowledge, R – retention, T – transfer, C – certainty, ^y – correct answers, ⁿ – incorrect answers, SE – self-evaluation

As in the previous case, in order to ensure the validity of our comparisons, we conducted assumption tests (Appendix 7). It is worth noting that no significant violations were observed, affirming the reliability of the subsequent analyses. As a result, Student's *t*-tests were employed to explore the differences between the groups and draw meaningful conclusions from the data.

Table 34 displays the comparison results, mirroring the outcomes of the immediate knowledge test. Even a week after the learning session, the two groups exhibited no statistically significant differences in their performance on the test, again showing no support for Hypothesis 6. Although studies frequently highlight the impact of an instructor's emotions on immediate test results, our results are consistent with prior research that typically fails to show differences in learning scores a week after viewing videos with an instructor displaying different emotions (Lawson et al., 2021a, 2021c; Lawson and Mayer, 2022). Additionally, there were no noteworthy variations in the level of certainty expressed in their answers or in their self-perception of test performance. These consistent findings further reinforce the notion that the type of narrator voice prosody (enthusiastic or calm) alone may not exert a significant influence on the participants' learning outcomes or their confidence in their test responses.

Table 34: Comparison between the enthusiastic and calm voice groups on various learning variables in the delayed part of the experiment using Student's *t*-tests

	<i>t</i>	<i>df</i>	<i>p</i>	<i>Mean difference</i>	<i>95% CI</i>	<i>d</i>	<i>95% CI</i>
K	10.791	92	0.283	0.894	−0.75–2.54	0.223	−0.19–0.63
R	11.030	92	0.273	0.638	−0.51–1.79	0.228	−0.18–0.63
T	0.730	92	0.467	0.255	−0.44–0.95	0.151	−0.26–0.56
C	0.191	94	0.849	0.725	−6.82–8.27	0.039	−0.36–0.44
C ^y	0.310	94	0.758	12.19	−6.60–9.04	0.063	−0.34–0.46
C ⁿ	0.114	94	0.910	0.432	−7.13–7.99	0.023	−0.38–0.42
RC	0.126	94	0.900	0.473	−6.98–7.93	0.026	−0.38–0.43
RC ^y	0.174	94	0.862	0.682	−7.11–8.48	0.036	−0.37–0.44
RC ⁿ	−0.016	94	0.988	−0.058	−7.50–7.39	−0.003	−0.40–0.40
TC	0.291	94	0.772	12.057	−7.02–9.44	0.059	−0.34–0.46
TC ^y	0.512	93	0.610	23.244	−6.69–11.34	0.105	−0.30–0.51
TC ⁿ	0.325	94	0.746	13.851	−7.07–9.84	0.067	−0.34–0.47
SE	−0.277	92	0.782	−0.064	−0.52–0.39	−0.057	−0.46–0.35

Note. *CI* – confidence interval, *d* – effect size; K – knowledge, R – retention, T – transfer, C – certainty, ^y – correct answers, ⁿ – incorrect answers, SE – self-evaluation

As was done previously, differences between groups were also tested by controlling for six confounding variables (prior interest, prior knowledge, English proficiency, and

affective state before watching the videos). Similar to the immediate testing phase, we applied a MANCOVA with seven dependent variables (retention score, transfer score, level of certainty in correct and incorrect retention answers, level of certainty in correct and incorrect transfer answers, and self-evaluated test performance) for the delayed testing phase, producing no significant results (Wilks' Lambda = 0.939, $F(7, 52) = 0.483$, $p = 0.843$; $\chi^2(28) = 20.701$, $p = 0.838$, $W = 0.875$, $p < .001$).

Regarding univariate tests, as all assumptions were met, we proceeded with ANCOVAs for all variables (Table 35). After controlling for covariates, the results remained unchanged and no significant differences were observed, so no post-hoc tests were made.

Table 35: ANCOVA comparisons with six covariates of the enthusiastic and calm voice groups on various learning variables in the delayed part of the experiment

	ANCOVA*			Homogeneity test**		Normality test	
	<i>F</i>	<i>p</i>	η^2p	<i>F</i>	<i>p</i>	<i>W</i>	<i>p</i>
Knowledge	1.794	0.184	0.020	0.828	0.365	0.982	0.231
Retention	1.586	0.211	0.018	0.008	0.930	0.985	0.375
Transfer	1.098	0.298	0.013	0.130	0.719	0.993	0.894
Certainty	0.145	0.704	0.002	0.754	0.387	0.989	0.606
Certainty in correct answers	0.183	0.670	0.002	0.337	0.563	0.989	0.580
Certainty in incorrect answers	0.111	0.740	0.001	0.606	0.438	0.992	0.831
R Certainty	0.111	0.740	0.001	1.190	0.278	0.989	0.615
R Certainty in correct answers	0.135	0.714	0.002	0.431	0.513	0.991	0.793
R Certainty in incorrect answers	0.022	0.882	0.000	0.641	0.425	0.988	0.549
T Certainty	0.195	0.660	0.002	1.388	0.242	0.991	0.778
T Certainty in correct answers	0.214	0.645	0.002	0.470	0.495	0.990	0.688
T Certainty in incorrect answers	0.326	0.570	0.004	0.759	0.386	0.992	0.806
Self-evaluation	0.102	0.750	0.001	0.120	0.730	0.964	0.011

Note. * $df_1 = 1$, $df_2 = 88$; ** $df_1 = 1$, $df_2 = 94$; R – retention, T – transfer

Comparison of objective and subjective test performance between sessions

In order to observe the behavior of the learning variables over time, we conducted pairwise Student's *t*-tests to compare the same variables in both immediate and delayed conditions. There was a moderate correlation between the two test scores ($r_{\text{knowledge}} = 0.601$, $p_{\text{knowledge}} < .001$; $r_{\text{retention}} = 0.570$, $p_{\text{retention}} < .001$; $r_{\text{transfer}} = 0.408$, $p_{\text{transfer}} < .001$) and subjective evaluations of test performance ($r_{\text{self-evaluation}} = 0.610$, $p_{\text{self-evaluation}} < .001$).

Descriptive statistics are presented in Table 36, and the comparisons between the learning variables, along with the results of the normality tests, are shown in Table 37.

Table 36: Descriptive statistics of the learning outcomes from the immediate ($N = 224$) and delayed part of the experiment ($N = 94$)

	<i>M</i>		<i>SD</i>		Min–Max		Skewness		Kurtosis	
	Imm	Del	Imm	Del	Imm	Del	Imm	Del	Imm	Del
K	13.46	12.09	4.50	4.02	5–26	4–27	0.62	0.51	–0.01	1.27
R	8.90	7.94	3.15	2.81	2–17	2–18	0.50	0.55	–0.20	1.04
T	4.56	4.15	1.97	1.69	0–10	0–9	0.15	0.14	–0.21	–0.10
C	52.03	43.01	20.10	18.51	0.34– 95.17	0– 79.31	–0.27	–0.18	–0.42	–0.36
C ^y	55.16	45.12	21.42	19.19	0– 95.91	0– 82.56	–0.28	–0.19	–0.61	–0.39
C ⁿ	48.44	41.36	19.16	18.55	0– 93.25	0– 80.36	–0.23	–0.11	–0.14	–0.37
RC	51.84	40.99	20.34	18.28	0.53– 97.37	0– 76.32	–0.22	–0.12	–0.52	–0.45
RC _y	55.27	43.45	21.90	19.12	0– 100	0– 83.67	–0.21	–0.07	–0.73	–0.42
RC _n	47.64	38.74	19.23	18.26	0–91	0– 77.78	–0.19	–0.01	–0.29	–0.49
TC	52.37	46.84	21.66	20.19	0– 92.60	0–85	–0.27	–0.17	–0.50	–0.43
TC _y	54.78	47.49	23.91	22.01	0– 100	0– 93.75	–0.21	–0.01	–0.75	–0.64
TC _n	50.18	45.62	22.00	20.74	0– 100	0–90	–0.04	–0.03	–0.28	–0.46
SE	3.33	3.05	1.17	1.11	1–7	1–5	–0.11	–0.40	0.49	–0.90

Note. Imm – immediate part of the experiment, Del – delayed part of the experiment; K – knowledge, R – retention, T – transfer, C – certainty, ^y – correct answers, ⁿ – incorrect answers, SE – self-evaluation

The assumption of normality was violated only when comparing subjective self-evaluations of test performance ($W = 0.88$, $p < .001$). Looking at the results (Table 37), there were mostly no significant differences in learning outcomes between the immediate and delayed session, except in the case of certainty in ones answers on retention questions, where the average level of certainty fell from 50.53% in the immediate session ($SD = 18.51$) to 43.01% ($SD = 18.51$) in the delayed session and, more specifically, certainty levels in answers that were answered correctly, falling from 53.33 ($SD = 19.51$) to 45.12 ($SD = 19.19$).

Table 37: Pairwise comparisons of the learning variables in the immediate and delayed parts of the experiment with normality test

Variable	<i>t</i>	<i>p</i>	<i>Mean difference</i>	<i>95% CI</i>	<i>d</i>	<i>95% CI</i>
Knowledge*	1.08	0.281	0.38	−0.32–1.08	0.11	−0.09–0.31
Retention*	1.66	0.101	0.44	−0.09–0.96	0.17	−0.03–0.37
Transfer*	−0.28	0.777	−0.05	−0.43–0.32	−0.03	−0.23–0.17
Certainty [†]	2.74	0.007	7.53	2.08–12.97	0.28	0.08–0.48
Certainty ^{y†}	2.93	0.004	8.21	2.65–13.77	0.30	0.09–0.50
Certainty ^{n†}	2.35	0.021	6.4	0.99–11.81	0.24	0.04–0.44
R Certainty [†]	3.32	0.001	9.14	3.68–14.59	0.34	0.13–0.54
R Certainty ^{y†}	3.52	< .001	10.02	4.37–15.67	0.36	0.15–0.56
R Certainty ^{n†}	2.97	0.004	8.12	2.69–13.55	0.30	0.10–0.51
T Certainty [†]	1.53	0.130	4.47	−1.33–10.28	0.16	−0.05–0.36
T Certainty ^{y‡}	1.98	0.051	5.94	−0.02–11.91	0.20	−0.00–0.41
T Certainty ^{n‡}	1.26	0.212	3.84	−2.22–9.90	0.13	−0.07–0.33
Self-evaluation*	2.48	0.015	0.25	0.05–0.44	0.26	0.05–0.46

Note. **df* = 93, [†]*df* = 95, [‡]*df* = 94; R – retention, T – transfer, ^y – correct answers, ⁿ – incorrect answers

Comparisons based on English proficiency

Since one of the main motivations for the study was to explore how people learn from multimedia in their non-native language, we were also interested in whether there will be differences in the results when examining participants with differing levels of English proficiency. As such, this section aims to replicate the previous comparisons but with a new twist: we will carry out these comparisons separately for individuals with lower and higher English proficiency. To categorize participants, we have chosen to use a LexTALE test score of 63 as the threshold. This decision is grounded in the fact that 51.77% of our sample (equivalent to 117 participants in the main part of the experiment and 55 participants in the delayed part; $M = 55.44$, $SD = 5.26$, $min = 37.50$, $max = 62.50$) scored below this threshold, while 48.23% (comprising 109 participants in the main part of the experiment and 43 in the delayed part; $M = 74.13$, $Mdn = 71.25$, $SD = 9.10$, $min = 63.75$, $max = 100.00$) scored above it. While the full set of results of ANCOVAs (or Quade's tests when assumptions were not met) is available in Appendices (8 to 11), only results of post-hoc tests from variables with significant main effects will be reported in the following part.

Lower proficiency group

Descriptive statistics and results of ANCOVAs, together with assumptions tests, for the lower English proficiency group, are available as tables in Appendices 8 and 9. When comparing only the participants that had a LexTALE score of less than 63, there were significant differences in the perception of the narrator's enthusiasm ($t(109) = 4.25$, $p < .001$, mean difference = 1.35, $d = 0.80$, 95% CI [0.41–1.19], $M_{ent} = 3.56$, $Mdn = 4.00$, $SD_{ent} = 1.69$, $M_{calm} = 2.26$, $Mdn = 1.00$, $SD_{calm} = 1.62$), calmness ($t(109) = -3.49$, $p < .001$, mean difference = -0.82, $d = -0.66$, 95% CI [-1.04—0.27], $M_{ent} = 5.32$, $Mdn = 5.00$, $SD_{ent} = 1.36$, $M_{calm} = 6.17$, $Mdn = 7.00$, $SD_{calm} = 1.20$), boredom ($t(109) = -3.33$, $p = 0.001$, mean difference = -1.13, $d = -0.63$, 95% CI [-1.01—0.25], $M_{ent} = 3.02$, $Mdn = 3.00$, $SD_{ent} = 1.77$, $M_{calm} = 4.12$, $Mdn = 4.00$, $SD_{calm} = 1.96$), pleasantness ($t(109) = 2.32$, $p = 0.022$, mean difference = 0.63, $d = 0.44$, 95% CI [0.06–0.82], $M_{ent} = 4.63$, $Mdn = 5.00$, $SD_{ent} = 1.46$, $M_{calm} = 4.09$, $Mdn = 4.00$, $SD_{calm} = 1.53$), and activation level ($t(109) = 2.43$, $p = 0.017$, mean difference = -0.59, $d = 0.46$, 95% CI [0.08–0.84], $M_{ent} = 3.76$, $Mdn = 4.00$, $SD_{ent} = 1.32$, $M_{calm} = 3.29$, $Mdn = 3.00$, $SD_{calm} = 1.38$).

In addition, there were also significant differences in perceiving the instructor as facilitating learning ($t(114) = 1.82$, $p = 0.071$, mean difference = 0.40, $d = 0.39$, 95% CI [0.01–0.76], $M_{ent} = 4.21$, $Mdn = 4.10$, $SD_{ent} = 0.87$, $M_{calm} = 3.85$, $Mdn = 3.90$, $SD_{calm} =$

1.20), credible ($t(114) = 2.03$, $p = 0.044$, mean difference = 0.53, $d = 0.52$, 95% CI [0.14–0.90], $M_{ent} = 5.29$, $Mdn = 5.20$, $SD_{ent} = 0.85$, $M_{calm} = 4.84$, $Mdn = 5.20$, $SD_{calm} = 1.25$), human-like ($t(109) = 2.08$, $p = 0.040$, mean difference = 0.55, $d = 0.39$, 95% CI [0.01–0.77], $M_{ent} = 4.44$, $Mdn = 3.87$, $SD_{ent} = 1.38$, $M_{calm} = 3.87$, $Mdn = 3.90$, $SD_{calm} = 1.42$), and engaging ($t(109) = 1.98$, $p = 0.050$, mean difference = 0.52, $d = 0.37$, 95% CI [–0.01–0.75], $M_{ent} = 3.48$, $Mdn = 3.40$, $SD_{ent} = 1.32$, $M_{calm} = 3.01$, $Mdn = 2.80$, $SD_{calm} = 1.45$).

There are, however, new differences between the enthusiastic and calm groups that were not evident when looking at the whole sample. There was a marginally significant difference in the level of intrinsic cognitive load ($t(109) = 1.88$, $p = 0.062$, mean difference = 0.40, $d = 0.36$, 95% CI [–0.02–0.73]), with the group viewing videos with the enthusiastic narrator reporting higher levels ($M_{ent} = 4.38$, $Mdn = 4.50$, $SD_{ent} = 1.11$, $M_{calm} = 4.02$, $Mdn = 4.00$, $SD_{calm} = 1.21$). This is surprising as intrinsic cognitive load relates to the inherent complexity of the educational content and should not be affected by the design of the learning material (Sweller, 1994; Sweller et al., 2011), unlike extraneous and germane cognitive load. Nevertheless, it offers initial insight that the additional cues provided by an enthusiastic narrator might increase the cognitive load for learners with a lower command of the language in which they are learning. The same group also reported lower self-evaluated test performance at the end on of the immediate knowledge test ($t(114) = -2.13$, $p = 0.035$, mean difference = –0.31, $d = 0.31$, 95% CI [–0.68–0.71], $M_{ent} = 3.14$, $Mdn = 3.00$, $SD_{ent} = 0.94$, $M_{calm} = 3.50$, $Mdn = 4.00$, $SD_{calm} = 1.22$). The final set of differences emerged in the delayed part of the experiment. A week after the initial viewing, participants with lower English proficiency showed significant differences between the groups in their knowledge ($t(47) = 2.88$, $p = 0.006$, mean difference = 2.95, $d = 0.81$, 95% CI [0.22–1.41], $M_{ent} = 12.58$, $Mdn = 12.50$, $SD_{ent} = 3.92$, $M_{calm} = 10.34$, $Mdn = 11.00$, $SD_{calm} = 3.55$), retention ($t(47) = 2.14$, $p = 0.038$, mean difference = 1.63, $d = 0.60$, 95% CI [0.02–1.18], $M_{ent} = 8.15$, $Mdn = 8.00$, $SD_{ent} = 2.95$, $M_{calm} = 6.86$, $Mdn = 7.00$, $SD_{calm} = 2.45$), and transfer levels ($t(47) = 2.93$, $p = 0.005$, mean difference = 1.31, $d = 0.83$, 95% CI [0.23–1.42], $M_{ent} = 4.42$, $Mdn = 4.50$, $SD_{ent} = 1.84$, $M_{calm} = 3.48$, $Mdn = 3.00$, $SD_{calm} = 1.53$), with the group with the enthusiastic narrator scoring higher than the group with the calm narrator. This aligns with previous literature on the impact of an instructor's emotional tone on learning (Lawson et al., 2021a, 2021b, 2021c) and demonstrates that narrators can positively influence learners simply by expressing enthusiasm through vocal cues. However, this effect appears to be effective only for learners who are less proficient in the language of the video. This might

also explain why the effect was not observed in the study by Davis et al. (2019), as their sample included English majors, indicating higher language proficiency.

In summary, among participants with lower English proficiency, those who viewed videos with an enthusiastic narrator perceived the instructor as significantly more facilitative, credible, human-like, engaging, enthusiastic, less calm, less boring, more pleasant, and more activated. These results did not differ from our main findings, encompassing all participants. What is new is that participants with lower English proficiency and the enthusiastic narrator also reported slightly higher levels of intrinsic cognitive load, self-evaluated their test performance as lower, yet performed significantly better in both retention and transfer a week after watching the videos, with medium and large effect sizes, giving partial support to Hypotheses 5 (regarding the difference in cognitive load) and 6 (regarding the difference in learning outcomes). These findings indicate that the same emotional design intervention may affect people with differing levels of language proficiency differently, as these results were not apparent when looking at the whole sample.

Higher proficiency group

Tables in Appendices 10 and 11 present the descriptive statistics, ANCOVA results (or Quade's tests), and assumption tests for the group with higher English proficiency. When exclusively analysing participants with a LexTALE score above 63, the initial results were similar to the those with a lower proficiency. There were (marginally) significant differences in the perception of the narrator's level of enthusiasm ($t(101) = 4.95, p < .001$, mean difference = 1.40, $d = 1.00$, 95% CI [0.57–1.42], $M_{ent} = 3.52$, $Mdn_{ent} = 4.00$, $SD_{ent} = 1.54$, $M_{calm} = 1.98$, $Mdn_{calm} = 1.00$, $SD_{calm} = 1.32$), calmness ($t(101) = -1.79, p = 0.076$, mean difference = -0.47 , $d = -0.36$, 95% CI [-0.77 – -0.04], $M_{ent} = 5.38$, $Mdn_{ent} = 6.00$, $SD_{ent} = 1.25$, $M_{calm} = 5.84$, $Mdn_{calm} = 6.00$, $SD_{calm} = 1.32$), boredom ($t(101) = -3.23, p = 0.002$, mean difference = -1.15 , $d = -0.65$, 95% CI [-1.07 – -0.24], $M_{ent} = 3.35$, $Mdn_{ent} = 3.00$, $SD_{ent} = 1.71$, $M_{calm} = 4.54$, $Mdn_{calm} = 5.00$, $SD_{calm} = 1.81$), and activation level ($t(101) = 3.631, p < .001$, mean difference = 1.06, $d = 0.73$, 95% CI [0.32–1.15], $M_{ent} = 3.81$, $Mdn_{ent} = 4.00$, $SD_{ent} = 1.47$, $M_{calm} = 2.74$, $Mdn_{calm} = 3.00$, $SD_{calm} = 1.54$), but not pleasantness. The two groups also differed significantly in their perception of the narrator's ability to facilitate learning ($t(101) = 2.07, p = 0.041$, mean difference = 0.46, $d = 0.42$, 95% CI [0.01–0.82], $M_{ent} = 4.25$, $Mdn_{ent} = 4.40$, $SD_{ent} = 0.95$, $M_{calm} = 3.72$, $Mdn_{calm} = 3.60$, $SD_{calm} = 1.24$), being engaging ($t(101) = 2.15, p = 0.034$, mean difference = 0.57, $d = 0.44$, 95% CI [0.03–0.84], $M_{ent} = 3.14$, $Mdn_{ent} = 3.10$, $SD_{ent} = 1.29$, $M_{calm} =$

2.43, $Mdn_{calm} = 2.20$, $SD_{calm} = 1.35$), their credibility ($t(101) = 6.01$, $p = 0.016$, mean difference = 0.51, $d = 0.54$, 95% CI [0.13–0.94], $M_{ent} = 5.26$, $Mdn_{ent} = 5.40$, $SD_{ent} = 0.74$, $M_{calm} = 4.76$, $Mdn_{calm} = 4.80$, $SD_{calm} = 1.39$), and human-likeness ($t(101) = 3.04$, $p = 0.003$, mean difference = 0.81, $d = 0.61$, 95% CI [0.20–1.02], $M_{ent} = 2.93$, $Mdn_{ent} = 4.10$, $SD_{ent} = 1.39$, $M_{calm} = 2.93$, $Mdn_{calm} = 2.60$, $SD_{calm} = 1.33$).

In contrast to the lower English proficiency group, notable variations were observed between the enthusiastic and calm narrator groups in the immediate knowledge test ($t(99) = -2.10$, $p = 0.038$, mean difference = 1.57, $d = -0.43$, 95% CI [-0.84–0.02]), particularly in the transfer segment ($t(99) = -2.50$, $p = 0.014$, mean difference = -0.77, $d = -0.51$, 95% CI [-0.93–0.10]). In this case, the results were opposite to those in the lower proficiency group – participants who viewed videos with the enthusiastic narrator scored lower ($M_{knowledge} = 13.81$, $Mdn_{knowledge} = 13.50$, $SD_{knowledge} = 4.39$, $M_{transfer} = 4.58$, $Mdn_{transfer} = 5.00$, $SD_{transfer} = 1.89$) compared to participants who watched the videos with the calm narrator ($M_{knowledge} = 15.55$, $Mdn_{knowledge} = 15.00$, $SD_{knowledge} = 5.16$, $M_{transfer} = 5.42$, $Mdn_{transfer} = 5.00$, $SD_{transfer} = 1.89$), with effect sizes indicating a small to medium negative impact. The trend persisted a week later, with those who had the enthusiastic narrator continuing to score lower on the knowledge test, and the effect size indicating a medium to large negative impact, albeit with marginal significance ($t(31) = -1.76$, $p = 0.089$, mean difference = -0.93, $d = -0.65$, 95% CI [-1.41–0.12], $M_{ent} = 4.10$, $Mdn_{ent} = 4.00$, $SD_{ent} = 1.26$, $M_{calm} = 4.89$, $Mdn_{calm} = 5.00$, $SD_{calm} = 1.88$). However, the same group had a higher level of certainty in their correct answers over the whole test ($t(35) = 1.84$, $p = 0.075$, mean difference = 10.24, $d = 0.63$, 95% CI [-0.08–1.33], $M_{ent} = 50.81$, $Mdn_{ent} = 50.00$, $SD_{ent} = 19.51$, $M_{calm} = 41.74$, $Mdn_{calm} = 41.59$, $SD_{calm} = 12.49$) and in every answer on the transfer part of the test ($t(35) = 2.959$, $p = 0.093$, mean difference = 10.38, $d = 0.61$, 95% CI [-0.10–1.32], $M_{ent} = 53.01$, $Mdn_{ent} = 48.50$, $SD_{ent} = 20.49$, $M_{calm} = 44.07$, $Mdn_{calm} = 43.00$, $SD_{calm} = 12.21$).

To sum up, on top of the differences regarding narrator perception that were apparent in the whole sample, the group with higher language proficiency and the enthusiastic narrator performed worse in the immediate and partly in the delayed knowledge test, especially the transfer part, but were more confident in their answers. These results again partly support Hypothesis 6, in which we predicted significant differences in learning outcomes between the groups with a calm and enthusiastic narrator. While the additional prosodic and emotional cues provided by the enthusiastic narrator's voice seem to benefit learners with lower English proficiency, they also seem to hinder more complex learning for those with higher English proficiency. This is a novel

finding, as it is the first study to compare non-native speakers' learning from videos with narrators expressing different emotions using voice alone, and to divide participants based on their language proficiency level. While the study by Davis et al. (2019) did not find differences in instructor perception, cognitive load, and learning between non-native students learning from weak and strong prosodic voices, their sample included students with a higher command of English. Our findings suggest a more complex connection, warranting further research to replicate the results and to investigate more thoroughly the elements that lead to such outcomes.

3.3.3.3 *Same-language subtitles*

The upcoming subsection will examine the comparison between participants who viewed the videos with and without same-language subtitles (SLS). While there are fewer hypotheses involved in this comparison compared to the previous section, which focused on different narrator emotional tones (namely, there are no hypotheses regarding instructor perception and emotional outcomes), the structure of the subsection will remain consistent. It will begin with results related to instructor perception, followed by findings on emotional, cognitive, and learning outcomes. This includes all variables measured in the experiment, along with the results of assumption tests.

Instructor perception

This section is divided into two parts, focusing on the initial two steps of the cognitive-affective model of e-learning: recognizing the emotion of the narrator and fostering a stronger social connection with the narrator.

Recognizing the emotion from the voice of the narrator

Table 38 presents the descriptive statistics of participant ratings for the video narrator, assessing their enthusiasm, calmness, nervousness, boredom, emotional valence, and activation level.

Table 38: Descriptive statistics for variables related to recognizing the narrator's emotion for groups with and without SLS

	<i>M</i>		<i>SD</i>		Min–Max		Skewness		Kurtosis	
	No SLS	SLS	No SLS	SLS	No SLS	SLS	No SLS	SLS	No SLS	SLS
ENTH	2.94	2.69	1.75	1.64	1–7	1–7	0.37	0.51	–1.00	0.88
CAL	5.68	5.69	1.33	1.33	1–7	2–7	–0.87	–0.73	0.26	–0.25
FRU	1.53	1.57	0.98	1.13	1–6	1–6	2.04	2.13	4.13	3.98
BOR	3.67	3.86	1.89	1.93	1–7	1–7	0.27	0.14	–0.85	–1.01
PL	4.32	4.45	1.49	1.39	1–7	1–7	–0.23	–0.30	–0.54	–0.28
AL	3.52	3.26	1.51	1.44	1–7	1–7	0.05	0.16	–0.56	–0.66

Note. No SLS – group without same-language subtitles, SLS – group with same-language subtitles; ENTH – enthusiasm, CAL – calmness, FRU – frustration, BOR – boredom, PL – pleasantness, AL – activation level

Prior to comparing the two groups in terms of ratings, we performed Levene's homogeneity test and Shapiro-Wilk normality test, while also inspecting box plots for outliers. The results indicated that both groups exhibited equal variances across all ratings. However, it should be noted that all variables violated the assumption of normality. As done before, despite this violation, we proceeded with Student's *t*-tests as the large sample size is likely to have minimal impact on the results.

Table 39: Comparison of the groups with and without SLS on discrete emotions, pleasantness and activation level rating of the narrator using *t*-tests

	<i>t</i>	<i>p</i>	<i>Mean difference</i>	<i>95% CI</i>	<i>d</i>	<i>95% CI</i>
ENTH	1.086	0.279	0.245	–0.20–0.69	0.145	–0.12–0.41
CALM	–0.087	0.930	–0.015	–0.36–0.33	–0.012	–0.27–0.25
FRU	–0.265	0.791	–0.037	–0.31–0.24	–0.035	–0.30–0.23
BOR	–0.734	0.464	–0.186	–0.69–0.31	–0.098	–0.36–0.16
PL	–0.670	0.503	–0.129	–0.51–0.25	–0.089	–0.35–0.17
AL	1.323	0.187	0.260	–0.13–0.65	0.176	–0.09–0.44

Note. *df* = 224, *CI* – confidence interval, *d* – effect size; ENTH – enthusiasm, CALM – calmness, FRU – frustration, BOR – boredom, PL – pleasantness, AL – activation level

As observed in Table 39, there were no significant differences in any of the emotion recognition variables between the group with and without SLS, which aligns with our expectations.

As was the case in the previous subchapter, the groups with and without SLS will also be compared while accounting for covariates. For this reason, a MANCOVA and ANCOVAs were performed adding the following covariates: prior interest in the topic, prior knowledge, English proficiency, and initial emotional state (as measured with PANAVA-KS).

As expected, the MANCOVA showed that even after including covariates there was no significant effect of SLS on perception of narrator's emotion (Wilks' Lambda = 0.947, $F(6, 213) = 0.985$, $p = 0.437$; $\chi^2(10) = 29.655$, $p = 0.099$, $W = 0.900$, $p < .001$). Table 40 presents the outcomes of univariate ANCOVAs, accompanied by assessments of the assumptions. As there were no significant differences observed, no further analyses were made.

Table 40: ANCOVA comparisons with six covariates of the groups without and with SLS on discrete emotions, pleasantness and activation level rating of the narrator

	ANCOVA*			Homogeneity test**		Normality test	
	<i>F</i>	<i>p</i>	η^2p	<i>F</i>	<i>p</i>	<i>W</i>	<i>p</i>
Enthusiasm	0.586	0.445	0.003	0.634	0.427	0.935	< .001
Calmness	0.148	0.700	0.001	0.051	0.822	0.929	< .001
Frustration	0.028	0.868	0.000	0.224	0.637	0.798	< .001
Boredom	0.111	0.740	0.001	0.040	0.841	0.974	< .001
Pleasantness	1.552	0.214	0.007	0.363	0.548	0.992	0.281
Activation level	0.714	0.399	0.003	0.034	0.854	0.990	0.102

Note. * $df_1 = 1$, $df_2 = 218$; ** $df_1 = 1$, $df_2 = 224$

Social partnership with the narrator

Similar to our predictions in the first step of the cognitive-affective model of e-learning, where we examined the recognition of the narrator's emotion, we anticipated no significant differences between the two groups in the second step of the model. This step was measured using the four components of API-R. A comprehensive overview of the descriptive statistics for this comparison is presented in Table 41.

Table 41: Descriptive statistics for API components for group without and with SLS

	<i>M</i>		<i>SD</i>		Min–Max		Skewness		Kurtosis	
	No SLS	SLS	No SLS	SLS	No SLS	SLS	No SLS	SLS	No SLS	SLS
FL	4.09	3.92	1.14	1.05	1.00–6.20	1.20–6.60	–0.57	–0.09	0.11	–0.20
CR	5.13	4.94	1.02	1.08	1.40–7.00	1.00–7.00	–0.86	–0.95	1.24	1.72
HL	3.96	3.60	1.49	1.44	1.00–6.80	1.00–6.80	–0.37	0.05	–0.75	–0.84
EN	3.14	2.89	1.50	1.28	1.00–7.00	1.00–6.80	0.33	0.46	–0.74	–0.14

Note. FL – facilitating learning, CR – credibility, HL – human-likeness, EN – engaging

All of the API components satisfied the assumption of equal variances; however, all of them did violate the assumption of normality. Nonetheless, we proceeded with Student's *t*-tests.

Table 42: Comparison of the groups without and with SLS on the Facilitating learning, Human-likeness, Credibility, and Engaging variables using *t*-tests

	<i>t</i>	<i>df</i>	<i>p</i>	<i>Mean difference</i>	<i>95% CI</i>	<i>d</i>	<i>95% CI</i>
Facilitating learning	1.154	224	0.250	0.168	–0.12–0.46	0.154	–0.11–0.42
Credibility	1.348	224	0.179	0.188	–0.09–0.46	0.179	–0.08–0.44
Human-likeness	1.851	224	0.065	0.362	–0.02–0.75	0.246	–0.02–0.51
Engaging	1.342	224	0.181	0.249	–0.12–0.62	0.179	–0.08–0.44

Note. *CI* – confidence interval, *d* – effect size

The results obtained from this part of the study and presented in Table 42 align with our expectations, indicating that the presence or absence of same-language subtitles (SLS) did not significantly impact the participants' perception of the narrator, as measured by the four components of API-R.

When comparing the two groups with the inclusion of covariates, the results (Table 43) were the same. A MANCOVA did not reveal any significant impact of including SLS on the outcome (Wilks' Lambda = 0.980, $F(4, 215) = 1.077$, $p < .001$; $\chi^2(10) = 9.407$, $p =$

0.494, $W = 0.959$, $p < .001$) and Table 43 displays the results of multiple ANCOVAs, again failing to find any significant differences.

Table 43: ANCOVA comparisons with five covariates of the groups without and with SLS on the Facilitating learning, Human-likeness, Credibility, and Engaging variables

	ANCOVA*			Homogeneity test**		Normality test	
	F	p	η^2p	F	p	W	p
Facilitating learning	0.326	0.569	0.001	0.016	0.900	0.991	0.170
Credibility	0.334	0.564	0.002	0.265	0.607	0.962	< .001
Human-like	2.518	0.114	0.011	0.073	0.788	0.985	0.017
Engaging	1.200	0.275	0.005	2.512	0.114	0.980	0.003

Note. * $df_1 = 1$, $df_2 = 218$; ** $df_1 = 1$, $df_2 = 224$

Emotional outcomes

This section contains various outcomes, including differences in affective states as measured by PANAVA-KS and the valence and activation level questions, interest in the topic, intrinsic motivation to watch the videos, and learners' experience.

Differences in affective state

As in the previous comparison between participants who viewed videos with an enthusiastic and calm narrator, this subsection will be divided into three parts, as the affective states were evaluated through three scales: PANAVA-KS, activation level, and valence scales. The first scale was administered before and after watching the videos, and the activation level and valence single-item scales were administered six times.

Starting with the PANAVA-KS, descriptive statistics are displayed in Table 44.

Table 44: Descriptive statistics of PANAVA-KS values and change score for groups without and with SLS

	<i>M</i>		<i>SD</i>		Min–Max		Skewness		Kurtosis	
	No SLS	SLS	No SLS	SLS	No SLS	SLS	No SLS	SLS	No SLS	SLS
PA ^b	3.62	3.46	1.00	1.19	1.25– 7.00	1.00– 6.75	0.32	0.09	0.68	– 0.19
PA ¹	3.11	2.91	1.20	1.16	1.00– 7.00	1.00– 5.75	0.48	0.23	0.25	– 0.43
PA ^c	–0.51	–0.55	1.13	1.08	–4.50– 2.25	–4.50– –2.00	– 0.26	– 0.93	1.28	1.94
NA ^b	3.22	3.33	1.16	1.24	1.00– 5.75	1.00– 6.50	0.03	0.24	– 0.67	– 0.41
NA ¹	3.03	3.06	1.01	1.11	1.00– 5.00	1.00– 6.25	– 0.33	0.10	– 0.59	– 0.17
NA ^c	–0.19	–0.27	1.02	0.90	–3.25– 2.25	–2.75– –1.50	– 0.42	– 0.66	0.55	0.34
VA ^b	4.62	4.54	1.06	1.26	1.50– 7.00	1.50– 7.00	– 0.02	– 0.25	0.03	– 0.38
VA ¹	4.45	4.24	0.98	1.15	2.50– 7.00	1.00– 6.50	0.12	– 0.42	– 0.39	0.24
VA ^c	–0.17	–0.30	1.18	1.11	–4.00– 3.00	–4.00– –2.50	– 0.20	– 0.40	0.73	0.87

Note. PA – positive activation, NA – negative activation, VA – valence, ^b – baseline, ¹ – measure after videos, ^c – change score (baseline measure subtracted from the measure after videos)

To investigate differences in the PANAVA-KS subscales between the groups without and with same-language subtitles (SLS), paired samples *t*-tests were conducted separately for each group. The assumption of normality was examined for the PANAVA-KS subscales first.

In the group without SLS, the positive activation, negative activation, and valence variables approached significance levels but still maintained a normal distribution after applying the Bonferroni correction. The normality assumption was met for these variables. Conversely, in the group with SLS, all PANAVA-KS subscales violated the assumption of normality. Despite this violation, due to the sample size, we proceeded

with parametric paired samples *t*-tests to analyse the differences in the PANAVA-KS subscales. The results of these tests can be found in Table 45.

Table 45: Baseline and post-intervention differences in the PANAVA-KS measures for the group without SLS and the group with SLS separately

	<i>t</i>	<i>df</i>	<i>p</i>	Mean difference	95% CI	<i>d</i>	95% CI
No same-language subtitles group							
PA	4.825	114	< .001	0.509	0.30–0.72	0.450	0.26–0.64
NA	2.028	114	0.045	0.193	0.00–0.38	0.189	0.00–0.37
NA	1.587	114	0.115	0.174	–0.04–0.39	0.148	–0.04–0.33
Same-language subtitles group							
PA	5.339	110	< .001	0.550	0.35–0.75	0.507	0.31–0.70
NA	3.131	110	0.002	0.268	0.10–0.44	0.297	0.11–0.49
VA	2.815	110	0.006	0.297	0.09–0.51	0.267	0.08–0.46

Note. CI – confidence interval, *d* – effect size; PA – positive activation, NA – negative activation, VA – valence

In both groups, participants experienced a significant decrease in positive activation after watching the videos, indicating a reduction in their overall positive emotional state. The effect size of this decrease was medium, suggesting a noticeable impact of the videos on participants' positive emotional response. Additionally, in the group with same-language subtitles (SLS), there was a significant decrease in negative activation after watching the videos compared to the baseline measure. This suggests that the videos had a positive influence on reducing negative emotional states among participants in this group. Furthermore, in the same group with SLS, there was a trend towards a lower valence score, approaching statistical significance. This suggests that participants' overall emotional valence became less positive after watching the videos with SLS, although it did not reach full statistical significance.

Upon analysing the affective states results for each group individually, we proceeded with three ANCOVAs to investigate the differences between the groups with and without same-language subtitles (SLS). In these ANCOVAs, the baseline measures on all three affective subscales were used as covariates to account for any initial differences in emotional states between the groups.

Assumption tests for normality and homogeneity were performed for the analysis. For positive activation, both assumptions were met ($W = 0.993$, $p = 0.383$; $F(1, 224) = 0.618$, $p = 0.433$), as well as for negative activation ($W = 0.988$, $p = 0.048$; $F(1, 224) = 0.653$, $p = 0.420$) and valence scales ($W = 0.989$, $p = 0.073$; $F(1, 224) = 0.001$, $p = 0.977$).

As anticipated, no significant differences were found between the two groups in positive activation ($F(1, 223) = 0.648$, $p = 0.422$, $\eta^2p = 0.003$), negative activation ($F(1, 223) = 0.062$, $p = 0.803$, $\eta^2p = 0.000$), and valence ($F(1, 223) = 1.888$, $p = 0.171$, $\eta^2p = 0.008$). Consequently, no post-hoc tests were conducted due to these findings.

Additionally, MANCOVA and ANCOVAs were used to test for differences between the two groups while also controlling for covariates. Although it appears that neither of the MANCOVA assumptions were met in this instance ($\chi^2(6) = 13.913$, $p = 0.031$, $W = 0.952$, $p < .001$), as previously stated, significant Shapiro-Wilk normality test and Box's test do not pose a problem in large and equivalent samples (Field, 2018), so we can proceed with MANCOVA. The results of MANCOVA indicated that SLS inclusion did not significantly affect the affective state of participants (Wilks' Lambda = 0.982, $F(3, 216) = 1.293$, $p = 0.278$).

In regards to ANCOVAs, no serious violations of assumptions was detected in any of the variables (positive activation scale: $W = 0.993$, $p = 0.329$; $F(1, 224) = 0.188$, $p = 0.665$; negative activation scale: $W = 0.988$, $p = 0.065$; $F(1, 224) = 0.458$, $p = 0.499$; valence scale: $W = 0.987$, $p = 0.043$; $F(1, 224) = 0.202$, $p = 0.654$). Even after including confounding variables, there was no significant difference between the two groups in positive activation ($F(1, 218) = 0.521$, $p = 0.471$, $\eta^2p = 0.002$), negative activation ($F(1, 218) = 0.144$, $p = 0.705$, $\eta^2p = 0.001$), and valence ($F(1, 218) = 0.622$, $p = 0.431$, $\eta^2p = 0.003$), so no post-hoc tests were performed.

In the subsequent subsection, the outcomes obtained from the activation level and valence single-item scales will be discussed. Both were administered both before and after watching each of the five videos. Descriptive statistics pertaining to these scales are provided in Table 46.

Table 46: Descriptive statistics of activation level and valence measurements and change score for groups without and with SLS

	<i>M</i>		<i>SD</i>		Min–Max		Skewness		Kurtosis	
	No SLS	SLS	No SLS	SLS	No SLS	SLS	No SLS	SLS	No SLS	SLS
AL ^b	4.77	4.60	1.57	1.55	1–9	1–7	–0.14	–0.26	–0.11	–0.57
AL ¹	4.92	4.76	1.58	1.46	1–9	2–8	–0.43	–0.19	0.23	–0.68
AL ²	4.63	4.42	1.57	1.51	1–8	1–7	–0.10	–0.21	–0.24	–0.68
AL ³	4.26	4.10	1.66	1.60	1–8	1–7	–0.16	–0.11	–0.51	–0.77
AL ⁴	4.37	4.08	1.76	1.54	1–9	1–7	–0.29	–0.12	–0.33	–0.47
AL ⁵	4.46	3.96	1.83	1.65	1–9	1–7	–0.07	–0.15	–0.27	–0.68
AL _M	4.53	4.26	1.46	1.31	1.00–7.80	1.40–7.20	–0.40	–0.10	–0.33	–0.57
AL _{cs}	–0.23	–0.34	1.58	1.35	–4.40–4.20	–4.40–3.80	0.17	–0.22	0.98	0.67
V ^b	5.34	5.50	1.50	1.52	2–9	2–8	0.06	–0.11	–0.67	–0.88
V ¹	5.32	5.52	1.32	1.27	1–8	2–8	–0.50	–0.06	0.70	–0.26
V ²	4.96	5.04	1.40	1.35	1–8	1–8	–0.39	–0.13	–0.15	0.14
V ³	4.63	4.74	1.45	1.51	1–8	1–8	–0.31	–0.51	–0.02	0.06
V ⁴	4.85	4.66	1.61	1.48	1–8	1–8	–0.51	–0.30	0.08	–0.11
V ⁵	4.77	4.60	1.67	1.63	1–9	1–8	–0.25	–0.59	0.30	0.04
V ^M	4.91	4.91	1.30	1.28	1.20–8.00	1.40–8.00	–0.49	–0.27	0.32	–0.04
V _{cs}	–0.43	–0.59	1.53	1.56	–4.40–3.60	–5.60–3.00	–0.11	–0.44	0.58	0.19

Note. AL – activation level, V – valence, ^b – baseline, ^M – average of the five responses after watching each video, ^{cs} – change score (baseline measure subtracted from the average score)

In a similar manner as before, paired samples *t*-tests were conducted separately in both groups to assess the differences from baseline in the various measures. Most variables violated the assumption of normality in both the group with and without SLS.

Table 47: Baseline and post-intervention differences in the activation level and valence measures for groups without and with SLS

	<i>t</i>	<i>p</i>	<i>Mean difference</i>	<i>95% CI</i>	<i>d</i>	<i>95% CI</i>
No same-language subtitles group*						
Activation I. ¹	-1.047	0.297	-0.157	-0.45–0.14	-0.098	-0.28–0.09
Activation I. ²	0.885	0.378	0.130	-0.16–0.42	0.082	-0.10–0.27
Activation I. ³	3.024	0.003	0.504	0.17–0.84	0.282	0.10–0.47
Activation I. ⁴	2.147	0.034	0.391	0.03–0.75	0.200	0.02–0.38
Activation I. ⁵	1.649	0.102	0.304	-0.06–0.67	0.154	-0.03–0.34
Activation I. ^M	1.598	0.113	0.235	-0.06–0.53	0.149	-0.04–0.3
Valence ¹	0.140	0.889	0.017	-0.23–0.26	0.013	-0.17–0.20
Valence ²	2.567	0.012	0.383	0.09–0.68	0.239	0.05–0.42
Valence ³	4.553	< .001	0.713	0.40–1.02	0.425	0.23–0.62
Valence ⁴	2.725	0.007	0.487	0.13–0.84	0.254	0.07–0.44
Valence ⁵	3.187	0.002	0.565	0.21–0.92	0.297	0.11–0.48
Valence ^M	3.032	0.003	0.433	0.15–0.72	0.283	0.10–0.47
Same-language subtitles group**						
Activation I. ¹	-1.240	0.218	-0.153	-0.40–0.09	-0.118	-0.30–0.07
Activation I. ²	1.220	0.225	0.180	-0.11–0.47	0.116	-0.07–0.30
Activation I. ³	3.309	0.001	0.505	0.20–0.81	0.314	0.12–0.50
Activation I. ⁴	3.239	0.002	0.523	0.20–0.84	0.307	0.12–0.50
Activation I. ⁵	3.881	< .001	0.640	0.31–0.97	0.368	0.18–0.56
Activation I. ^M	2.638	0.010	0.339	0.08–0.59	0.250	0.06–0.44
Valence ¹	-0.135	0.892	-0.018	-0.28–0.25	-0.013	-0.20–0.17
Valence ²	3.164	0.002	0.468	0.18–0.76	0.300	0.11–0.49
Valence ³	4.498	< .001	0.766	0.43–1.10	0.427	0.23–0.62
Valence ⁴	5.057	< .001	0.847	0.52–1.18	0.480	0.28–0.68

Valence ⁵	4.881	< .001	0.901	0.54–1.27	0.463	0.27–0.66
Valence ^M	4.011	< .001	0.593	0.30–0.89	0.381	0.19–0.57

Note. * $df_1 = 114$, ** $df_2 = 110$, CI – confidence interval, d – effect size, ^M - average

The results presented in Table 47 reveal some changes in the valence and activation level measures between the baseline and post-intervention assessments in both groups. These changes were even more prominent in the group with SLS and demonstrated a small to medium effect size. This indicates that after watching the videos, participants in the group with SLS experienced a slight decrease in activation levels and participants in both groups experienced a small decrease in pleasantness compared to their baseline levels.

Subsequently, we conducted ANCOVAs to examine potential differences between the groups with and without SLS, while also conducting Levene's and Shapiro-Wilk tests to assess assumptions violations. The results showed that the assumption of equal variances was met for all variables, but some variables did violate the normality assumption. Despite this, we proceeded with ANCOVAs, using baseline measures as covariates. The outcomes of the ANCOVAs and assumptions checks can be found in Table 48.

As anticipated, no significant differences were observed between the two groups, which is why post-hoc tests were not conducted. The lack of significant differences suggests that the presence of same-language subtitles did not significantly impact the valence and activation level measures or lead to divergent outcomes in comparison to the group without SLS.

Table 48: ANCOVA comparisons of the groups with and without SLS on activation level and valence items

Subscale	ANCOVA*			Homogeneity test**		Normality test	
	F	p	η^2p	F	p	W	p
Activation level ¹	0.211	0.647	0.001	2.493	0.116	0.978	0.001
Activation level ²	0.549	0.460	0.002	0.027	0.869	0.992	0.231
Activation level ³	0.204	0.652	0.001	0.474	0.492	0.986	0.028
Activation level ⁴	1.259	0.263	0.006	2.097	0.149	0.990	0.133
Activation level ⁵	3.943	0.048	0.017	2.179	0.141	0.994	0.471
Activation level ^M	1.452	0.229	0.006	1.991	0.160	0.987	0.034

Valence ¹	0.727	0.395	0.003	1.597	0.208	0.980	0.003
Valence ²	0.012	0.914	0.000	0.014	0.906	0.981	0.004
Valence ³	0.103	0.749	0.000	0.147	0.701	0.972	< .001
Valence ⁴	1.468	0.227	0.007	0.482	0.488	0.973	< .001
Valence ⁵	1.048	0.307	0.005	0.000	0.998	0.971	< .001
Valence ^M	0.105	0.746	0.000	0.172	0.679	0.974	< .001

Note. * $df_1 = 1$, $df_2 = 223$; ** $df_1 = 1$, $df_2 = 224$; ^M – average

As anticipated, no significant differences were observed between the two groups, which is why post-hoc tests were not conducted. However, when a MANCOVA was conducted with the ten singular measurements of valence and activation level as outcomes and prior interest, prior knowledge, English proficiency, and both baseline measures of valence and activation level as covariates, a marginally significant effect emerged (Wilks' Lambda = 0.920, $F(10, 210) = 0.454$, $p = 0.059$; $\chi^2(55) = 107.245$, $p < .001$, $W = 0.817$, $p < .001$). Conversely, multiple follow up ANCOVAs presented in Table 49 did not reveal a significant effect. The lack of significant differences suggests that the presence of same-language subtitles did not significantly impact the valence and activation level measures or lead to divergent outcomes in comparison to the group without SLS.

Table 49: ANCOVA comparisons with five covariates of the groups with and without SLS on activation level and valence items

	ANCOVA*			Homogeneity test**		Normality test	
	<i>F</i>	<i>p</i>	η^2p	<i>F</i>	<i>p</i>	<i>W</i>	<i>p</i>
Activation level ¹	0.266	0.606	0.001	2.682	0.103	0.982	0.005
Activation level ²	0.954	0.330	0.004	0.002	0.969	0.995	0.680
Activation level ³	0.156	0.693	0.001	0.255	0.614	0.987	0.042
Activation level ⁴	1.035	0.310	0.005	1.600	0.207	0.992	0.227
Activation level ⁵	3.294	0.071	0.015	1.595	0.208	0.996	0.764
Activation level ^M	1.409	0.237	0.006	1.625	0.204	0.987	0.041
Valence ¹	2.301	0.131	0.010	0.339	0.561	0.993	0.419
Valence ²	0.391	0.533	0.002	0.002	0.961	0.987	0.033
Valence ³	0.903	0.343	0.004	0.108	0.743	0.985	0.018

Valence ⁴	0.330	0.566	0.002	0.809	0.369	0.980	0.003
Valence ⁵	0.173	0.678	0.001	0.098	0.754	0.981	0.005
Valence ^M	0.146	0.703	0.001	0.056	0.814	0.978	0.001

Note. * $df_1 = 1$, $df_2 = 219$; ** $df_1 = 1$, $df_2 = 224$; ^M – average

Interest in the topic

Interest in the topic was assessed using two ways: a short questionnaire in the first session and a question asking participants to what extent they were interested in the topic before beginning to answer test questions in the second session.

Regarding the situational interest questionnaire, no assumptions were violated. The results showed that participants who watched the videos without SLS did not significantly differ in their level of interest in the topic of the videos ($t(223) = 1.681$, $p = 0.094$; $M = 3.58$, $SD = 1.11$) compared to participants who viewed the videos with added SLS ($M = 3.32$, $SD = 1.25$). The same result can be observed when accounting for prior interest, prior knowledge, English proficiency, and initial emotional state ($F(1, 217) = 0.980$, $p = 0.323$, $\eta^2p = 0.004$; $W = 0.995$, $p = 0.734$; $F(1, 223) = 3.467$, $p = 0.064$).

Secondly, one week after the initial viewing, 94 participants rated, on a single item, how interesting they found the topic of the videos. Since the assumption of equal variances was approaching significance, a Welch's t -test was conducted to account for potentially heterogeneous variances. Similarly to the previous results, there were no significant differences in the answer to the question between the two groups ($t(86.114) = 2.152$, $p = 0.034$; $M_{no\ SLS} = 3.85$, $SD_{no\ SLS} = 1.09$; $M_{SLS} = 3.30$, $SD_{SLS} = 1.36$). Analysis of covariance revealed the same result ($F(1, 86) = 1.487$, $p = 0.226$, $\eta^2p = 0.017$; $W = 0.992$, $p = 0.874$; $F(1, 92) = 0.619$, $p = 0.433$). These findings indicate that the presence of same-language subtitles did not significantly influence the immediate and delayed level of interest in the topic for participants in the study.

Intrinsic motivation

The assumption of normality approached significance, but the assumption of equal variances was met, allowing the use of a Student's t -test to compare whether the addition of SLS affects participants' motivation. As anticipated, the results revealed no significant differences ($t(223) = 1.183$, $p = 0.238$; $M_{no\ SLS} = 3.54$, $SD_{no\ SLS} = 1.19$; $M_{SLS} = 3.35$, $SD_{SLS} = 1.25$), indicating that the presence of SLS did not have a significant impact on participants' motivation. Similar results were obtained even when controlling for

covariates ($F(1, 217) = 0.239, p = 0.625, \eta^2p = 0.001; W = 0.995, p = 0.640; F(1, 223) = 0.568, p = 0.452$).

Learners' experience

In the following section, the results of a set of five questions that are commonly used in multimedia learning research to gauge learners' experience with the learning videos are presented. Participants were asked about their motivation to pay attention, the perceived difficulty of the lectures, the amount of effort they exerted to learn the information, their overall enjoyment of the experience, and whether they would be interested in viewing similar lessons in the future. Descriptive statistics for these questions can be found in Table 50.

Table 50: Descriptive statistics for the learners' experience questions for groups without and with SLS

	<i>M</i>		<i>SD</i>		Min–Max		Skewness		Kurtosis	
	No SLS	SLS	No SLS	SLS	No SLS	SLS	No SLS	SLS	No SLS	SLS
PA	3.34	3.38	1.37	1.53	1–6	1–7	0.26	0.10	–0.51	–0.98
DIF	3.52	3.59	1.40	1.47	1–7	1–7	0.11	0.24	–0.57	–0.58
EF	3.61	3.53	1.48	1.38	1–7	1–7	–0.05	–0.04	–1.01	–0.47
ENJ	3.68	3.55	1.42	1.52	1–7	1–7	0.12	–0.19	–0.45	–0.84
ML	3.52	3.07	1.61	1.47	1–7	1–6	0.12	0.03	–0.71	–0.86

Note. Ent – enthusiastic narrator group, Cal – calm narrator group; PA – paying attention, DIF – difficulty, EF – exerting more effort, ENJ – enjoyment, ML – more lessons like this

All variables met the assumption of equal variances, but did not meet the normality assumption, allowing us to perform Student's *t*-tests.

Table 51: Comparison of the groups without and with SLS on several variables on the learners' experience

	<i>t</i>	<i>p</i>	<i>Mean difference</i>	<i>95% CI</i>	<i>d</i>	<i>95% CI</i>
Paying attention	-0.187	0.852	-0.036	-0.42–0.35	-0.025	-0.29–0.24
Difficulty	-0.403	0.688	-0.077	-0.45–0.30	-0.054	-0.32–0.21
Exerting more effort	0.431	0.667	0.083	-0.29–0.46	0.058	-0.20–0.32
Enjoyment	0.687	0.493	0.135	-0.25–0.52	0.092	-0.17–0.35
More lessons like this	2.163	0.032	0.445	0.04–0.85	0.288	0.02–0.55

Note. *df* = 223, *CI* – confidence interval, *d* – effect size

The results displayed in Table 51 reveal that there were no significant differences between the two groups of students who watched the videos with or without SLS, except in the case of the last item. This indicates that the addition of SLS did not have a significant impact on learners' experience with the videos in terms of their motivation to pay attention, perceived difficulty of the lectures, effort exerted to learn the information, and overall enjoyment of the experience. However, there was a marginally significant difference between the two groups in their wish for similar lessons in the future, with those who watched videos without the additional SLS reporting they want more lessons like this more compared to participants who viewed videos with SLS.

The same findings emerged when adding prior knowledge, prior interest, English proficiency, and initial affective state (PANAVA subscales) as confounding variables. According to MANCOVA, the inclusion of SLS produced marginally significant impact on the learners' experience variables (Wilks' Lambda = 0.957, $F(5, 213) = 1.896$, $p = 0.096$; $\chi^2(15) = 23.260$, $p = 0.079$, $W = 0.954$, $p < .001$).

Additionally, Table 52 presents the outcomes of the multiple ANCOVAs designed to assess whether the inclusion of covariates had any impact on the results. ANCOVA results reveal the same trend - the difference in wanting more similar lessons remained marginally significant ($t(217) = 1.710$, $p = 0.089$, mean difference = 0.335, $d = 0.232$, 95% *CI* [-0.036–0.499]). However, this result should be interpreted with caution, especially in the absence of any other similar results. As can be observed from Table 52, the other results were not significant. The variable "Paying attention" was analysed

with the non-parametric Quade test due to the assumption of homogeneity not being met, also showing nonsignificant results ($F(1,223) = 0.636, p = .426$).

Table 52: ANCOVA comparisons with six covariates of the groups without and with SLS on several variables on the learners' experience

	ANCOVA*			Homogeneity test**		Normality test	
	<i>F</i>	<i>p</i>	η^2p	<i>F</i>	<i>p</i>	<i>W</i>	<i>p</i>
Paying attention				5.597	0.019	0.995	0.595
Difficulty	0.310	0.578	0.001	0.835	0.362	0.993	0.376
Exerting more effort	0.086	0.770	0.000	1.168	0.281	0.988	0.048
Enjoyment	0.007	0.934	0.000	0.298	0.585	0.997	0.904
More lessons like this	2.925	0.089	0.013	0.559	0.456	0.991	0.163

Note. * $df_1 = 1, df_2 = 217$; ** $df_1 = 1, df_2 = 223$

As was done in the subchapter on the narrator emotional tone, a MANCOVA was also performed with the variables "Situational interest," "Intrinsic motivation," and "Enjoyment" ($r > 0.770, p < .001$) as outcomes due to the high correlation between them seen in Appendix 3. MANCOVA did not reveal any significant impact (Wilks' Lambda = 0.977, $F(3, 215) = 1.675, p = 0.173$; $\chi^2(6) = 8.924, p = 0.178, W = 0.975, p < .001$).

Cognitive outcomes

Variables related to cognitive outcomes contain measures of perceived cognitive load and mental effort. Hypothesis 7 predicts that there will be a significant difference in the level of cognitive load between participants who watched the videos with SLS and those who did not.

Cognitive load

Table 53 presents descriptive statistics categorized by SLS addition group, distinguishing between intrinsic, extraneous, and germane cognitive load.

Table 53: Descriptive statistics of the cognitive load questionnaire for groups without and with SLS

	<i>M</i>		<i>SD</i>		Min–Max		Skewness		Kurtosis	
	No SLS	SLS	No SLS	SLS	No SLS	SLS	No SLS	SLS	No SLS	SLS
ICL	3.97	4.03	1.19	1.31	1.00–6.50	1.50–7.00	0.02	0.01	–0.61	–0.74
ECL	3.82	3.52	1.21	1.24	1.33–7.00	1.00–7.00	0.30	0.43	–0.45	0.05
GCL	4.61	4.41	0.97	1.25	2.00–7.00	1.00–7.00	–0.32	–0.50	–0.21	–0.26

Note. ICL – intrinsic cognitive load, ECL – extraneous cognitive load, GCL – germane cognitive load

The assumption of homogeneity of variances was not violated in either case, as evidenced by the results of the Levene's tests displayed in Appendix 7. However, the assumption of normality was violated for intrinsic and germane cognitive load. Independent Student's *t*-tests were performed to check for potential differences between groups. The results are presented in Table 54.

Table 54: Comparison of the groups without and with SLS on cognitive load

	<i>t</i>	<i>p</i>	<i>Mean difference</i>	<i>95% CI</i>	<i>d</i>	<i>95% CI</i>
ICL	–0.321	0.749	–0.053	–0.38–0.27	–0.043	–0.30–0.22
ECL	1.829	0.069	0.299	–0.02–0.62	0.244	–0.02–0.51
GCL	1.339	0.182	0.200	–0.09–0.49	0.179	–0.08–0.44

Note. *df* = 223, *CI* – confidence interval, *d* – effect size; ICL – intrinsic cognitive load, ECL – extraneous cognitive load, GCL – germane cognitive load

While there were no differences expected in the level of intrinsic and germane cognitive load, Hypothesis 7 predicted that there will be a difference in extraneous cognitive load between participants who watched the videos with the addition of SLS and those who watched the videos without SLS. Table 54 reveals that there was a difference in the level of extraneous cognitive load between the two groups, albeit with marginal significance ($p = 0.069$) and a small effect size. The findings indicate that participants without the help of SLS reported slightly higher extraneous cognitive load, which is in line with findings from some of the previous studies on using SLS when learning in a

foreign language (Lee and Mayer, 2018; Lin et al., 2016; Mayer et al., 2014), in which participants who learned from videos with narration and added text reported lower difficulty, effort, and cognitive load compared to learners who viewed video without subtitles. While our results should be interpreted with caution, they suggest that the inclusion of same-language subtitles in the videos may relieve some of the cognitive load while learning in our non-native language.

Additionally, MANCOVA with prior knowledge, interest, initial emotional state, and English proficiency as covariates and all three cognitive load measures as outcomes revealed a marginally significant effect (Wilks' Lambda = 0.959, $F(3, 215) = 3.029$, $p = 0.030$; $\chi^2(6) = 9.319$, $p = 0.156$, $W = 0.981$, $p = 0.003$).

Subsequent ANCOVAs (as shown in Table 55) and the Quade test yielded results consistent with the t -tests. There were no significant effects observed in intrinsic ($F(1,223) = 0.235$, $p = 0.615$) and germane cognitive load, but there was a marginally significant effect on extraneous cognitive load, with a reduced p -value of 0.035 ($t(217) = 2.118$, $p = 0.035$, mean difference = 0.3369, $d = 0.287$, 95% CI [0.019–0.555]).

Table 55: ANCOVA comparisons with six covariates of the groups without and with SLS on cognitive load

	ANCOVA*			Homogeneity test**		Normality test	
	F	p	η^2p	F	p	W	p
Intrinsic cognitive load				0.987	0.320	0.993	0.341
Extraneous cognitive l.	4.485	0.035	0.020	0.019	0.890	0.987	0.034
Germane cognitive load	0.740	0.391	0.003	3.491	0.063	0.981	0.004

Note. * $df_1 = 1$, $df_2 = 217$; ** $df_1 = 1$, $df_2 = 223$

Mental effort

Following each video, participants were asked to report the mental effort they invested in each of the videos, resulting in five separate measures of mental effort. Table 56 presents the descriptive statistics of these five measures, as well as their average.

Table 56: Descriptive statistics of the mental effort ratings for groups without and with SLS

	<i>M</i>		<i>SD</i>		Min–Max		Skewness		Kurtosis	
	No SLS	SLS	No SLS	SLS	No SLS	SLS	No SLS	SLS	No SLS	SLS
ME ¹	4.64	4.70	1.70	1.59	1–9	1–9	0.13	–0.23	–0.40	–0.50
ME ²	5.02	4.68	1.51	1.60	2–9	1–9	–0.03	–0.30	–0.23	–0.05
ME ³	4.58	4.60	1.79	1.63	1–9	1–9	0.02	–0.26	–0.35	–0.44
ME ⁴	4.57	4.48	1.69	1.65	1–9	1–9	–0.16	–0.51	0.10	–0.43
ME ⁵	4.71	4.36	1.78	1.66	1–9	1–9	–0.04	–0.64	0.16	–0.41
ME ^M	4.71	4.56	1.46	1.40	1.40– 8.80	1.20– 8.80	0.06	–0.47	0.38	–0.26

Note. ME – mental effort, ^M – average

The assumption of homogeneity of variance was satisfied for all variables, but the assumption of normality was violated for almost all variables.

Table 57: Comparison of the groups without and with SLS on mental effort

	<i>t</i>	<i>p</i>	<i>Mean difference</i>	<i>95% CI</i>	<i>d</i>	<i>95% CI</i>
Mental effort ¹	–0.270	0.787	–0.059	–0.49–0.37	–0.036	–0.30–0.22
Mental effort ²	1.654	0.100	0.342	–0.07–0.75	0.220	–0.04–0.48
Mental effort ³	–0.092	0.927	–0.021	–0.47–0.43	–0.012	–0.27–0.25
Mental effort ⁴	0.435	0.664	0.096	–0.34–0.53	0.058	–0.20–0.32
Mental effort ⁵	1.543	0.124	0.353	–0.10–0.80	0.205	–0.06–0.47
Mental effort ^M	0.746	0.456	0.142	–0.23–0.52	0.099	–0.16–0.36

Note. *df* = 224, *CI* – confidence interval, *d* – effect size; ^M – average

Table 57 presents a summary of the findings from the comparison between the two experimental groups concerning their levels of mental effort. The table includes data on all five mental effort items, as well as their average scores. No statistically significant differences were observed between the group watching videos with added SLS and the group without SLS. However, when all five mental effort measurements were taken together and covariates were added to the analysis, a MANCOVA revealed a marginally significant result (Wilks' Lambda = 0.936, $F(5, 214) = 2.927$, $p = 0.014$; $\chi^2(15) = 35.755$,

$p = 0.002$, $W = 0.892$, $p < .001$). Nevertheless, subsequent ANCOVAs (Table 58) failed to show any significant effect. These findings suggest that while there may be a subtle influence of SLS on mental effort when considering all variables together, this effect does not appear to be strong or consistent when examining individual mental effort measurements. Further research is needed to elucidate these results.

Table 58: ANCOVA comparisons with six covariates of the groups without and with SLS on mental effort

	ANCOVA*			Homogeneity test**		Normality test	
	F	p	η^2p	F	p	W	p
Mental effort ¹	0.294	0.588	0.001	0.570	0.451	0.995	0.632
Mental effort ²	2.616	0.107	0.012	0.603	0.438	0.992	0.276
Mental effort ³	0.104	0.747	0.000	0.430	0.513	0.985	0.017
Mental effort ⁴	0.023	0.879	0.000	0.398	0.529	0.976	< .001
Mental effort ⁵	2.323	0.129	0.011	0.759	0.385	0.983	0.007
Mental effort ^M	0.299	0.585	0.001	0.586	0.445	0.982	0.006

Note. * $df_1 = 1$, $df_2 = 218$; ** $df_1 = 1$, $df_2 = 224$; ^M – average

Learning outcomes

Similar to the section on narrator emotional tone, this part of the results will cover various objective and subjective learning-related outcomes. These outcomes include measures of retention, transfer, certainty in the correctness of participants' answers, and self-evaluated test performance, all categorized according to the immediate testing session ($N = 224$) and the delayed testing session ($N = 94$).

Objective and subjective test performance in the immediate part of the experiment

Table 59 presents the descriptive statistics for all learning-related outcomes during the main phase of the experiment. As previously noted, the variable "knowledge" represents all cumulative points earned on the test, while the "retention" and "transfer" variables consist of points accumulated from correctly answering questions related to either retention or transfer.

Table 59: Descriptive statistics of the learning outcomes from the immediate part of the experiment for groups without and with SLS

	<i>M</i>		<i>SD</i>		Min–Max		Skewness		Kurtosis	
	No SLS	SLS	No SLS	SLS	No SLS	SLS	No SLS	SLS	No SLS	SLS
K	13.15	13.77	3.98	4.99	5–26	5–26	0.90	0.40	1.09	–0.62
R	8.68	9.12	2.76	3.52	4–16	2–17	0.75	0.29	0.45	–0.62
T	4.46	4.65	1.89	2.06	0–10	0–9	0.17	0.12	0.31	–0.60
C	51.44	52.64	20.36	19.91	0.34– 91.72	3.45– 95.17	–0.28	–0.25	–0.29	–0.53
C ^y	55.44	54.88	21.87	21.05	0.00– 93.33	2.50– 95.91	–0.35	–0.21	–0.45	–0.78
C ⁿ	47.74	49.16	19.73	18.61	0.48– 93.25	0.00– 92.86	–0.17	–0.29	–0.18	–0.05
RC	51.30	52.40	20.75	19.99	0.53– 91.26	4.74– 97.37	–0.23	–0.20	–0.42	–0.62
RC ^y	55.62	54.91	22.20	21.68	0.00– 96.43	5.00– 100.0 0	–0.31	–0.11	–0.57	–0.87
RC ⁿ	47.34	47.94	20.07	18.40	0.83– 91.00	0.00– 87.50	–0.11	–0.29	–0.30	–0.27
TC	51.69	53.09	22.18	21.19	0.00– 92.60	0.00– 92.00	–0.26	–0.28	–0.46	–0.52
TC ^y	54.66	54.91	24.45	23.43	0.00– 100.0 0	0.00– 100.0 0	–0.18	–0.24	–0.69	–0.81
TC ⁿ	49.03	51.37	21.99	22.06	0.00– 100 0	0.00– 100.0 0	–0.19	0.12	–0.37	–0.21
SE	3.44	3.22	1.15	1.18	1–7	1–6	0.17	–0.37	1.13	–0.23

Note. K – knowledge, R – retention, T – transfer, C – certainty, ^y – correct answers, ⁿ – incorrect answers, SE – self-evaluation

The results of Shapiro-Wilk's tests of normality and Levene's tests of homogeneity of variances for the learning variables in the immediate part of the experiment represented in Appendix 7 show that the variables "Knowledge" and "Retention" not only violate the assumption of normality, but also equality of variances. Therefore, to ensure

accuracy and simplicity, Welch's test was conducted for all learning-related variables. This approach is justified as the results of Welch's and Student's *t*-tests are the same in cases where homogeneity of variances is met (Delacre et al., 2017).

Table 60: Comparison between the groups without and with SLS on various learning variables in the immediate part of the experiment using Welch's *t*-tests

	<i>t</i>	<i>df</i>	<i>p</i>	Mean difference	95% CI	<i>d</i>	95% CI
K	-1.032	208.17	0.303	-0.624	-1.81–0.57	-0.138	-0.40–0.12
R	-1.025	206.67	0.306	-0.434	-1.27–0.40	-0.137	-0.40–0.13
T	-0.718	218.90	0.474	-0.190	-0.71–0.33	-0.096	-0.36–0.17
C	-0.446	221.96	0.656	-1.201	-6.50–4.10	-0.060	-0.32–0.20
C ^y	0.194	222.00	0.847	0.555	-5.10–6.21	0.026	-0.24–0.29
C ⁿ	-0.553	221.89	0.581	-1.417	-6.47–3.63	-0.074	-0.34–0.19
RC	-0.403	222.00	0.688	-1.096	-6.46–4.27	-0.054	-0.32–0.21
RC ^y	0.240	221.97	0.810	0.705	-5.07–6.48	0.032	-0.23–0.29
RC ⁿ	-0.234	221.43	0.815	-0.602	-5.67–4.47	-0.031	-0.29–0.23
TC	-0.483	221.98	0.629	-1.400	-7.11–4.31	-0.065	-0.33–0.20
TC ^y	-0.080	219.99	0.936	-0.257	-6.59–6.08	-0.011	-0.27–0.25
TC ⁿ	-0.795	220.80	0.427	-2.346	-8.16–3.47	-0.107	-0.37–0.16
SE	1.411	221.13	0.160	0.220	-0.09–0.53	0.189	-0.08–0.45

Note. CI – confidence interval, *d* – effect size; K – knowledge, R – retention, T – transfer, C – certainty, ^y – correct answers, ⁿ – incorrect answers, SE – self-evaluation

Table 60 demonstrates that there were no statistically significant differences between the two groups in any of the learning outcomes. This lack of significance was also evident in the variables measuring participants' certainty in their answers and their subjective assessment of their displayed test performance.

Furthermore, a MANCOVA was conducted to account for prior interest, knowledge, initial emotional state, and English proficiency and with the following outcomes: retention and transfer test scores, level of certainty in correct answers, level of certainty in incorrect answers (both for retention and transfer parts of the test separately), and self-evaluated test performance. No significant effect was observed on this group of variables (Wilks' Lambda = 0.958, $F(7, 207) = 1.307$, $p = 0.248$; $\chi^2(28) = 32.794$, $p = 0.243$, $W = 0.884$, p

< .001). Additionally, similar findings were observed when the covariates were incorporated into single-variable analyses, such as ANCOVAs (see Table 61) and Quade tests. The latter analyses were performed because the overall ($F(1,222) = 1.159$, $p = 0.283$) and retention test results ($F(1,222) = 1.257$, $p = 0.263$) did not adhere to the assumption of homogeneity. In all cases, no statistically significant distinctions were detected. Therefore, the evidence does not lend support to Hypothesis 8 predicting significant differences in test results between those who viewed the educational videos with and without SLS. While the lack of differences in learning outcomes between the groups contradicts some of the previous studies (Lee and Mayer, 2018; Lin et al., 2016), the null results replicate the findings from other studies (Kraft, 2020; Liu et al., 2018; Pannatier and Béntrancourt, 2024; van der Zee et al., 2017), especially the study made by Mayer and his colleagues (2014), in which the authors found that while the inclusion of SLS alleviates perceived cognitive effort, the positive outcome does not transfer to better learning performance.

Table 61: ANCOVA comparisons with six covariates of the groups with and without SLS on various learning variables in the immediate part of the experiment

	ANCOVA*			Homogeneity test**		Normality test	
	<i>F</i>	<i>p</i>	η^2p	<i>F</i>	<i>p</i>	<i>W</i>	<i>p</i>
Knowledge				5.583	0.019	0.994	0.543
Retention				4.471	0.036	0.993	0.367
Transfer	0.646	0.422	0.003	1.536	0.216	0.996	0.805
Certainty	0.118	0.732	0.001	0.491	0.484	0.986	0.023
Certainty in correct answers	0.171	0.679	0.001	0.957	0.329	0.008	0.930
Certainty in incorrect answers	0.283	0.595	0.001	0.719	0.397	0.990	0.111
R Certainty	0.122	0.727	0.001	0.539	0.464	0.989	0.084
R Certainty in correct answers	0.172	0.679	0.001	0.605	0.437	0.986	0.023
R Certainty in incorrect answers	0.074	0.786	0.000	0.402	0.527	0.991	0.156
T Certainty	0.089	0.766	0.000	1.346	0.247	0.984	0.014
T Certainty in correct answers	0.038	0.845	0.000	1.302	0.255	0.992	0.227

T Certainty in incorrect answers	0.404	0.526	0.002	0.273	0.602	0.995	0.659
Self-evaluation	1.812	0.180	0.008	0.016	0.898	0.981	0.005

Note. * $df_1 = 1$, $df_2 = 216$; ** $df_1 = 1$, $df_2 = 222$; R – retention, T – transfer

Objective and subjective test performance in the delayed part of the experiment

Following the main experiment, a week later, 94 participants responded to the same test questions, provided ratings for their certainty in their answers, and subjectively assessed their test performance. 48 respondents were from the group with no subtitles and 46 from the group with subtitles. The results will be presented in the same manner as in the previous subsection, starting with descriptive statistics in Table 62.

Table 62: Descriptive statistics of the learning outcomes from the delayed part of the experiment for groups without and with SLS

	<i>M</i>		<i>SD</i>		Min–Max		Skewness		Kurtosis	
	No SLS	SLS	No SLS	SLS	No SLS	SLS	No SLS	SLS	No SLS	SLS
K	11.54	12.65	3.13	4.74	4–21	4–27	0.04	0.44	0.92	0.59
R	7.73	8.15	2.37	3.22	2–14	3–18	–0.01	0.69	0.28	0.79
T	3.81	4.50	1.23	2.02	2–7	0–9	0.23	–0.22	–0.42	–0.42
C	45.93	39.96	17.42	19.29	7.38– 75.62	0.00– 79.31	0.01	–0.26	–0.83	–0.22
C ^y	47.97	42.15	18.16	19.98	7.06– 82.56	0.00– 78.89	0.02	–0.29	–0.64	–0.41
C ⁿ	44.65	37.92	17.50	19.16	7.77– 75.00	0.00– 80.36	0.04	–0.15	–0.97	–0.10
RC	44.08	37.77	17.31	18.89	5.79– 73.21	0.00– 76.32	0.07	–0.19	–0.88	–0.33
RC ^y	45.91	40.88	17.89	20.19	7.82– 83.67	0.00– 78.89	0.19	–0.18	–0.44	–0.61
RC ⁿ	42.35	34.98	17.88	18.07	3.00– 72.22	0.00– 77.78	0.07	–0.08	–1.12	–0.00
TC	49.44	44.13	18.75	21.45	10.40 – 82.50	0.00– 85.00	–0.07	–0.15	–0.90	–0.24

TC ^y	51.23	43.50	21.85	21.70	5.40– 85.00	0.00– 93.75	0.02	–0.06	–1.18	–0.15
TC ⁿ	48.40	42.71	18.09	23.02	14.00 – 80.14	0.00– 90.00	–0.04	0.13	–0.80	–0.42
SE	3.08	3.02	1.07	1.16	1–5	1–5	–0.28	–0.49	–1.03	–0.84

Note. K – knowledge, R – retention, T – transfer, C – certainty, ^y – correct answers, ⁿ – incorrect answers, SE – self-evaluation

As before, to ensure the validity of our comparisons, assumption tests were performed. As one of the variables (Transfer) violated the assumption of equal variances, the comparison between the two groups was conducted using Welch's *t*-tests.

Given the anticipated modest impact of the added SLS, we also explored marginally significant findings. As demonstrated in Table 63, five marginally significant distinctions emerged – transfer, the degree of certainty in incorrect answers overall, and the level of certainty in responses to retention questions and incorrect answers specifically. Additionally, differences were observed in the level of certainty in accurate responses to transfer questions. The results of the transfer test seven days after the learning session were higher for the group with SLS, showing some limited support for Hypothesis 8, while the group without them expressed greater certainty in specific aspects of their responses.

Table 63: Comparison between the groups without and with SLS on various learning variables in the delayed part of the experiment using Welch's *t*-tests

	<i>t</i>	<i>df</i>	<i>p</i>	<i>Mean difference</i>	<i>95% CI</i>	<i>d</i>	<i>95% CI</i>
K	–1.334	77.60	0.186	–1.111	–2.77–0.55	–0.276	–0.68–0.13
R	–0.723	82.54	0.471	–0.423	–1.59–0.74	–0.150	–0.56–0.26
T	–1.983	73.83	0.051	–0.688	–1.38–0.01	–0.411	–0.82–0.00
C	1.589	92.10	0.115	5.970	–1.49–13.43	0.325	–0.08–0.73
C ^y	1.490	92.27	0.140	5.812	–1.94–13.56	0.304	–0.10–0.71
C ⁿ	1.795	92.37	0.076	6.733	–0.71–14.18	0.367	–0.04–0.77
RC	1.706	92.46	0.091	6.316	–1.04–13.67	0.349	–0.06–0.75
RC ^y	1.290	91.59	0.200	5.033	–2.71–12.78	0.264	–0.14–0.67
RC ⁿ	2.010	93.74	0.047	7.377	0.09–14.67	0.410	–0.00–0.82
TC	1.290	91.20	0.200	5.313	–2.87–13.50	0.264	–0.14–0.67

TC ^y	1.728	92.70	0.087	7.725	−1.15–16.60	0.355	−0.06–0.76
TC ⁿ	1.342	87.28	0.183	5.686	−2.74–14.11	0.275	−0.13–0.68
SE	0.267	90.51	0.790	0.062	−0.40–0.52	0.055	−0.35–0.46

Note. CI – confidence interval, *d* – effect size; K – knowledge, R – retention, T – transfer, C – certainty, ^y – correct answers, ⁿ – incorrect answers, SE – self-evaluation

Similar, marginally significant results emerged when accounting for multiple covariates with a MANCOVA and multiple ANCOVAs. As was the case when analysing the data from the immediate testing session, a MANCOVA was conducted with seven outcomes: retention score, transfer score, level of certainty in correct and incorrect retention responses, level of certainty in correct and incorrect transfer responses, and self-assessed test performance. The analysis yielded marginally significant results (Wilks' Lambda = 0.726, $F(7, 52) = 2.806$, $p = 0.015$; $\chi^2(28) = 36.383$, $p = 0.133$, $W = 0.875$, $p < .001$).

As was the case with the *t*-tests, ANCOVAs (see Table 64) revealed that marginally significant differences existed between the groups with and without SLS in various aspects, including overall test performance ($t(86) = -1.868$, $p = 0.065$, mean difference = -1.589 , $d = -0.425$, 95% CI [-0.881 – 0.032]), transfer test performance (Quade test; $F(1,92) = 8.425$, $p = 0.005$, $t(92) = -2.903$), the level of certainty in incorrect answers throughout the entire test ($t(86) = 1.812$, $p = 0.073$, mean difference = 7.533 , $d = 0.411$, 95% CI [-0.044 – 0.865]), and in the level of certainty in overall ($t(86) = 1.715$, $p = 0.090$, mean difference = 7.111 , $d = 0.389$, 95% CI [-0.066 – 0.843]) and incorrect answers during the retention section of the test ($t(86) = 1.993$, $p = 0.049$, mean difference = 8.192 , $d = 0.452$, 95% CI [-0.004 – 0.907]). After including confounding variables, the difference in overall test performance became marginally significant and had a small to medium effect, while the level of certainty in correct answers in the transfer parts of the test did not remain marginally significantly different. For the remaining variables, the *p*-values remained relatively consistent, except for transfer, where the *p*-value decreased. Again, the group who watched the videos with added SLS performed better overall and in the transfer part of the test, while those who watched the videos without SLS believed they performed better and expressed greater certainty in their responses. These findings suggest that the inclusion of SLS in educational videos could potentially enhance learners' performance on a transfer test conducted one week after the initial lesson. This is an important addition to the literature on the effectiveness of adding SLS, as all previous studies on the topic included only immediate testing, stressing the importance of verifying the long-term effect of SLS on learning (e.g., Lee and Mayer, 2018; Pannatier

and Bétrancourt, 2024). However, it's important to note that further research is necessary to substantiate this claim, as the observed results only reached marginal significance.

Table 64: ANCOVA comparisons with six covariates of the groups with and without SLS on various learning variables in the delayed part of the experiment

	ANCOVA*			Homogeneity test**		Normality test	
	<i>F</i>	<i>p</i>	η^2p	<i>F</i>	<i>p</i>	<i>W</i>	<i>p</i>
Knowledge	3.490	0.065	0.039	3.635	0.060	0.980	0.164
Retention	1.147	0.287	0.013	2.377	0.127	0.985	0.378
Transfer				7.701	0.007	0.990	0.685
Certainty	2.533	0.115	0.028	0.139	0.710	0.991	0.763
Certainty in correct answers	1.920	0.169	0.021	0.689	0.409	0.989	0.579
Certainty in incorrect answers	3.283	0.073	0.036	0.001	0.979	0.994	0.962
R Certainty	2.940	0.090	0.032	0.001	0.979	0.989	0.625
R Certainty in correct answers	1.698	0.196	0.019	1.074	0.303	0.989	0.625
R Certainty in incorrect answers	3.971	0.049	0.043	0.681	0.411	0.990	0.709
T Certainty	1.650	0.202	0.018	0.617	0.434	0.993	0.884
T Certainty in correct answers	1.945	0.167	0.022	0.010	0.920	0.987	0.504
T Certainty in incorrect answers	1.687	0.197	0.019	3.220	0.076	0.992	0.870
Self-evaluation	0.043	0.837	0.000	0.264	0.608	0.966	0.014

Note. * $df_1 = 1$, $df_2 = 88$; ** $df_1 = 1$, $df_2 = 94$; R – retention, T – transfer

Comparisons based on English proficiency

Similar to the analysis involving the emotional tone of the narrator, we conducted comparisons between the groups exposed to SLS and those without SLS, taking into consideration varying levels of tested English proficiency (as well as the following covariates: prior interest, prior tested knowledge, LexTALE score, and baseline measures of emotional state). Considering the level of language proficiency in studying

multimedia learning in a foreign language was also proposed by other authors (e.g., Mayer et al., 2014). One group comprised participants scoring below 63 on the LexTALE test, while the second group consisted of those who scored above 63. The detailed outcomes of all these comparisons are provided in Appendices 12 to 15. However, in the subsequent sections, we will only highlight and discuss significant findings.

Lower proficiency group

Descriptive statistics, ANCOVA (Quade's test) results, and the outcomes of assumption tests for the lower English proficiency group can be found in tables presented in Appendices 12 and 13. When focusing solely on participants who scored below 63 on the LexTALE test, there were marginally significant differences in five variables. Participants who viewed videos without SLS reported higher situational interest ($t(114) = 1.87, p = 0.064$, mean difference = 0.32, $d = 0.33$, 95% $CI [-0.05-0.71]$, $M_{no\ SLS} = 3.71$, $Mdn_{no\ SLS} = 3.83$, $SD_{no\ SLS} = 1.03$, $M_{SLS} = 3.20$, $Mdn_{SLS} = 3.25$, $SD_{SLS} = 1.26$) and a wish to have more lessons that are similar to the one they viewed ($t(109) = 1.77, p = 0.080$, mean difference = 0.51, $d = 0.34$, 95% $CI [-0.04-0.72]$, $M_{no\ SLS} = 3.69$, $Mdn_{no\ SLS} = 4.00$, $SD_{no\ SLS} = 1.66$, $M_{SLS} = 3.03$, $Mdn_{SLS} = 3.00$, $SD_{SLS} = 1.47$). A week after viewing the videos, the same group displayed a greater level of certainty in their answers when they were incorrect ($t(45) = 1.99, p = 0.053$, mean difference = 13.48, $d = 0.64$, 95% $CI [-0.02-1.30]$, $M_{no\ SLS} = 44.96$, $Mdn_{no\ SLS} = 47.38$, $SD_{no\ SLS} = 18.57$, $M_{SLS} = 35.22$, $Mdn_{SLS} = 38.13$, $SD_{SLS} = 22.05$) and in all answers on the retention part of the delayed test ($t(45) = 1.74, p = 0.089$, mean difference = 11.69, $d = 0.56$, 95% $CI [-0.10-1.22]$, $M_{no\ SLS} = 43.89$, $Mdn_{no\ SLS} = 39.74$, $SD_{no\ SLS} = 17.98$, $M_{SLS} = 35.67$, $Mdn_{SLS} = 40.00$, $SD_{SLS} = 21.76$). However, those who watched videos with SLS had a higher score on the delayed transfer test ($t(53) = -1.90, p = 0.063$, mean difference = -0.82, $d = -0.49$, 95% $CI [-1.08-0.11]$, $M_{no\ SLS} = 3.67$, $Mdn_{no\ SLS} = 4.00$, $SD_{no\ SLS} = 1.27$, $M_{SLS} = 4.18$, $Mdn_{SLS} = 4.50$, $SD_{SLS} = 2.07$) compared to those who did not watch videos with SLS, with an effect size indicating a medium impact.

Higher proficiency group

There were only two marginally significant differences in the group which scored higher than 63 on LexTALE. First, participants who had the aid of SLS viewed the narrator as more pleasant than those who did not ($t(101) = -1.89, p = 0.061$, mean difference = -0.46, $d = -0.37$, 95% $CI [-0.76-0.02]$, $M_{no\ SLS} = 4.18$, $Mdn_{no\ SLS} = 4.00$, $SD_{no\ SLS} = 1.38$, $M_{SLS} = 4.66$, $Mdn_{SLS} = 5.00$, $SD_{SLS} = 1.33$). On the other hand, respondents

who learned without SLS were more interested in the topic one week after the initial learning than those who watched videos with SLS ($t(31) = 1.79$, $p = 0.083$, mean difference = 0.97, $d = 0.76$, 95% $CI [-0.13-1.65]$, $M_{no\ SLS} = 4.05$, $Mdn_{no\ SLS} = 4.00$, $SD_{no\ SLS} = 1.20$, $M_{SLS} = 3.17$, $Mdn_{SLS} = 3.00$, $SD_{SLS} = 1.43$). Results of all other variables can be found in Appendices 14 and 15.

To sum up, when dividing the sample based on English proficiency, it becomes evident that the difference in delayed transfer scores between the groups is primarily due to learners with lower English proficiency. This indicates that SLS are more beneficial for long-term learning among learners with lower proficiency, rather than those with higher proficiency. Additionally, it was observed that viewers of videos without SLS found the video topic more interesting, though this perception varied across different variables for those with lower and higher proficiency. Both these results contradict the findings of a recent study that found no differences in learning performance, cognitive load, and situational interest, regardless of language proficiency (Pannatier and Bétrancourt, 2024).

3.3.3.4 Interactions

Although no specific hypotheses were formulated to predict an interaction between narrator emotion and the inclusion of same-language subtitles, we conducted multiple separate two-way ANCOVAs to explore any potential effects for the main dependent variables. The outcomes of these analyses are detailed in Tables 65 through 68, which include information on both individual effects and interactions.

Table 65: Two-way ANCOVA comparisons of the instructor perception variables, together with homogeneity tests

		ANCOVA*			Levene's test**	
		<i>F</i>	<i>p</i>	η^2p	<i>F</i>	<i>p</i>
Narrator affective state						
Enthusiasm	Narrator emotion	44.735	< .001	0.172	0.851	0.468
	SLS	0.843	0.359	0.004		
	Interaction	0.111	0.739	0.001		
Calmness	Narrator emotion	15.987	< .001	0.069	1.457	0.227
	SLS	0.191	0.663	0.001		
	Interaction	0.664	0.416	0.003		

Frustration	Narrator emotion	0.061	0.805	0.000	1.133	0.336
	SLS	0.024	0.877	0.000		
	Interaction	0.175	0.676	0.001		
Boredom	Narrator emotion	24.923	< .001	0.103	0.150	0.929
	SLS	0.162	0.688	0.001		
	Interaction	0.618	0.433	0.003		
Pleasantness	Narrator emotion	7.894	0.005	0.035	1.022	0.384
	SLS	1.448	0.230	0.007		
	Interaction	5.439	0.021	0.025		
Activation level	Narrator emotion	17.388	< .001	0.075	0.334	0.801
	SLS	0.843	0.360	0.004		
	Interaction	0.438	0.509	0.002		
Narrator perception						
Facilitating learning	Narrator emotion	9.173	0.003	0.041	2.257	0.083
	SLS	0.391	0.532	0.002		
	Interaction	0.139	0.710	0.001		
Credibility	Narrator emotion	13.136	< .001	0.057	5.191	0.002
	SLS	0.418	0.519	0.002		
	Interaction	0.171	0.680	0.001		
Human-like	Narrator emotion	16.227	< .001	0.070	1.237	0.297
	SLS	2.941	0.088	0.013		
	Interaction	1.471	0.226	0.007		
Engaging	Narrator emotion	10.390	0.001	0.046	0.893	0.446
	SLS	1.414	0.236	0.007		
	Interaction	2.403	0.123	0.011		

Note. * $df_1 = 1$, $df_2 = 216$; ** $df_1 = 3$, $df_2 = 222$

Among the various instructor perception variables analysed in this study, only one interaction between narrator emotion and SLS was found to be (marginally) statistically significant (Table 65) – perceived pleasantness of the instructor. Subsequent post-hoc comparisons revealed a significant difference in how participants perceived the pleasantness of the instructor when exposed to videos featuring enthusiastic versus calm narrators in the condition without added SLS ($t(216) = 3.653$, $p < .001$, $p_{\text{bonferroni}} = 0.002$, mean difference = 0.919, $d = 0.690$, 95% CI [0.312–1.069]). There was also a marginally significant difference between the group who watched videos with the enthusiastic narrator with SLS and the group watching videos with the calm narrator without SLS

($t(216) = 2.852$, $p = 0.005$, $p_{\text{bonferroni}} = 0.029$, mean difference = 0.719, $d = -0.540$, 95% CI [-0.540– -0.163]). The rest of comparisons were insignificant.

Table 66: Two-way ANCOVA comparisons of the emotional outcomes variables, together with homogeneity tests

		ANCOVA*			Levene's test**	
		<i>F</i>	<i>p</i>	η^2p	<i>F</i>	<i>p</i>
Participants' affective state						
Positive activation	Narrator emotion	1.941	0.165	0.009	0.343	0.794
	SLS	0.507	0.477	0.002		
	Interaction	0.262	0.609	0.001		
Negative activation	Narrator emotion	0.782	0.378	0.004	0.489	0.690
	SLS	0.129	0.720	0.001		
	Interaction	4.540	0.034	0.021		
Valence	Narrator emotion	1.530	0.218	0.007	0.516	0.672
	SLS	0.622	0.431	0.003		
	Interaction	0.933	0.335	0.004		
Activation level [†]						
	Narrator emotion	1.126	0.290	0.005	0.490	0.690
	SLS	1.388	0.240	0.006		
	Interaction	0.001	0.982	0.000		
Valence [†]						
	Narrator emotion	0.079	0.779	0.000	0.305	0.822
	SLS	0.141	0.707	0.001		
	Interaction	0.302	0.583	0.001		
Interest and motivation						
Situational interest	Narrator emotion	0.633	0.427	0.003	1.067	0.364
	SLS	1.002	0.318	0.005		
	Interaction	1.808	0.180	0.008		
Interest (delayed)	Narrator emotion	0.293	0.590	0.003	1.542	0.209
	SLS	1.366	0.246	0.016		
	Interaction	0.012	0.913	0.000		
Intrinsic motivation	Narrator emotion	1.643	0.201	0.008	0.718	0.542
	SLS	0.270	0.604	0.001		
	Interaction	1.216	0.271	0.006		
Learners' experience						
Paying attention	Narrator emotion	0.205	0.651	0.001	2.857	0.038

	SLS	0.639	0.425	0.003		
	Interaction	2.639	0.106	0.012		
Difficulty	Narrator emotion	1.121	0.291	0.005	0.407	0.748
	SLS	0.289	0.591	0.001		
	Interaction	0.239	0.626	0.001		
Exerting more effort	Narrator emotion	0.017	0.896	0.000	0.636	0.593
	SLS	0.097	0.756	0.000		
	Interaction	1.537	0.216	0.007		
Enjoyment	Narrator emotion	0.050	0.823	0.000	0.778	0.508
	SLS	0.008	0.927	0.000		
	Interaction	0.204	0.652	0.001		
More lessons like this	Narrator emotion	0.792	0.375	0.004	0.275	0.844
	SLS	3.024	0.083	0.014		
	Interaction	1.627	0.203	0.008		

Note. Participants' affective state (PANAVA-KS): * $df_1 = 1$, $df_2 = 216$; ** $df_1 = 3$, $df_2 = 222$; Activation level, Valence: * $df_1 = 1$, $df_2 = 217$; ** $df_1 = 3$, $df_2 = 222$; Situational interest, Intrinsic motivation, Learners' experience: * $df_1 = 1$, $df_2 = 215$; ** $df_1 = 3$, $df_2 = 221$; Interest (delayed): * $df_1 = 1$, $df_2 = 84$; ** $df_1 = 3$, $df_2 = 90$; †instead of using the PANAVA-KS baseline measurements, the activation level and valence baseline measurements were used

There were no statistically significant interactions found between narrator emotion and SLS concerning any of the emotional outcome variables at the p -value levels of 0.002 or 0.05 (Table 66). However, a marginal interaction effect was found for the negative activation, with subsequent post-hoc tests showing a marginally significant difference between the enthusiastic and calm group with SLS ($t(216) = 2.122$, $p = 0.035$, $p_{\text{bonferroni}} = 0.210$, mean difference = 0.327, $d = 0.405$, 95% CI [0.027–0.783]) that became insignificant if applying the Bonferroni correction.

Table 67: Two-way ANCOVA comparisons of the cognitive outcomes variables, together with homogeneity tests

			ANOVA*			Levene's test**	
			F	p	η^2p	F	p
Intrinsic cognitive load		Narrator emotion	0.389	0.534	0.002	2.119	0.099
		SLS	0.291	0.590	0.001		
		Interaction	0.460	0.498	0.002		
		Narrator emotion	1.977	0.161	0.009	0.998	0.395

Extraneous cognitive load	SLS	4.415	0.037	0.020	2.030	0.111
	Interaction	0.006	0.936	0.000		
	Narrator emotion	0.247	0.620	0.001		
Germane cognitive load	SLS	0.795	0.373	0.004	0.277	0.842
	Interaction	4.914	0.028	0.022		
	Narrator emotion	0.504	0.479	0.002		
Mental effort average	SLS	0.310	0.578	0.001	0.000	
	Interaction	0.020	0.888	0.000		
	Narrator emotion	0.504	0.479	0.002		

Note. Cognitive load: * $df_1 = 1$, $df_2 = 215$; ** $df_1 = 3$, $df_2 = 221$; Mental effort average: * $df_1 = 1$, $df_2 = 216$; ** $df_1 = 3$, $df_2 = 222$

As Table 67 shows, there are no significant interactions in the cognitive outcomes variables, except for germane cognitive load, which is marginally significant. Post-hoc tests revealed two marginally significant differences (enthusiastic narrator no SLS vs. enthusiastic narrator with SLS: ($t(215) = 2.180$, $p = 0.030$, $p_{\text{bonferroni}} = 0.182$, mean difference = 0.457, $d = 0.419$, 95% CI [0.038–0.800]); enthusiastic narrator with SLS vs. calm narrator with SLS: ($t(215) = -1.913$, $p = 0.057$, $p_{\text{bonferroni}} = 0.342$, mean difference = -0.398, $d = -0.365$, 95% CI [-0.743–0.013])) that became insignificant when applying the Bonferroni correction.

Table 68: Two-way ANCOVA comparisons of the main learning variables, together with homogeneity tests

		ANOVA*			Levene's test**	
		F	p	η^2p	F	p
Immediate testing						
Knowledge	Narrator emotion	1.765	0.185	0.008	2.274	0.081
	SLS	1.554	0.214	0.007		
	Interaction	0.097	0.755	0.000		
Retention	Narrator emotion	0.494	0.483	0.002	1.592	0.192
	SLS	1.350	0.247	0.006		
	Interaction	0.000	0.993	0.000		
Transfer	Narrator emotion	3.028	0.083	0.014	1.739	0.160
	SLS	0.686	0.409	0.003		
	Interaction	0.435	0.510	0.002		
Certainty	Narrator emotion	0.054	0.816	0.000	2.293	0.079
	SLS	0.111	0.739	0.001		

	Interaction	0.303	0.583	0.001		
Certainty in correct answers	Narrator emotion	0.024	0.878	0.000	2.572	0.055
	SLS	0.174	0.677	0.001		
	Interaction	0.707	0.401	0.003		
Certainty in incorrect answers	Narrator emotion	0.436	0.510	0.002	3.125	0.027
	SLS	0.267	0.606	0.001		
	Interaction	0.142	0.707	0.001		
Self-evaluation	Narrator emotion	3.327	0.070	0.015	0.055	0.983
	SLS	1.742	0.188	0.008		
	Interaction	0.027	0.869	0.000		
Delayed testing						
Knowledge	Narrator emotion	2.144	0.147	0.025	1.581	0.199
	SLS	3.646	0.060	0.042		
	Interaction	0.988	0.323	0.012		
Retention	Narrator emotion	1.744	0.190	0.020	0.513	0.674
	SLS	1.201	0.276	0.014		
	Interaction	1.145	0.288	0.013		
Transfer	Narrator emotion	1.508	0.223	0.018	2.463	0.068
	SLS	7.154	0.009	0.078		
	Interaction	0.287	0.594	0.003		
Certainty [†]	Narrator emotion	0.081	0.777	0.001	0.630	0.597
	SLS	2.267	0.136	0.026		
	Interaction	0.302	0.584	0.003		
Certainty in correct answers [†]	Narrator emotion	0.120	0.730	0.001	0.428	0.733
	SLS	1.654	0.202	0.019		
	Interaction	0.537	0.466	0.006		
Certainty in incorrect answers [†]	Narrator emotion	0.050	0.824	0.001	0.718	0.543
	SLS	2.989	0.087	0.034		
	Interaction	0.266	0.607	0.003		
Self-evaluation	Narrator emotion	0.093	0.761	0.001	2.280	0.085
	SLS	0.035	0.852	0.000		
	Interaction	0.000	0.994	0.000		

Note. Immediate testing: * $df_1 = 1$, $df_2 = 214$; ** $df_1 = 3$, $df_2 = 220$; Delayed testing: * $df_1 = 1$, $df_2 = 84$; ** $df_1 = 3$, $df_2 = 90$; [†] * $df_1 = 1$, $df_2 = 86$; ** $df_1 = 3$, $df_2 = 92$

Table 68 displays the findings from several two-way ANCOVAs, including learning variables assessed in both the immediate and delayed testing sessions with no noteworthy interactions. Two of the delayed testing variables did not satisfy the homogeneity of variances assumption, but none of the variables exhibited a significant interaction effect, so no inferences can be made.

Comparisons based on English proficiency

Following the structure of the previous results, two-way ANCOVAs were also performed separately for the groups with lower and higher English proficiency. The full set of results can be seen in Appendices 16 and 17, while in this section, only significant and marginally significant interactions will be highlighted.

Lower proficiency group

In the lower English proficiency group, there were three marginally significant interactions – narrator pleasantness, perceiving the narrator as engaging, and delayed transfer (Appendix 16).

There was a significant difference in narrator pleasantness in those who watched videos with an enthusiastic narrator without SLS and those with a calm narrator, both without SLS ($t(107) = 3.119$, $p = 0.002$, $p_{\text{bonferroni}} = 0.014$, mean difference = 1.195, $d = 0.840$, 95% CI [0.294–1.386]) and with SLS ($t(107) = 1.782$, $p = 0.078$, $p_{\text{bonferroni}} = 0.465$, mean difference = 0.702, $d = 0.493$, 95% CI [–0.059–1.046]). However, only the first difference remained significant after applying Bonferroni's correction.

In the case of perceiving the narrator as engaging, there were (marginally) significant differences between the group who watched videos with an enthusiastic narrator without SLS and the group who watched the videos with an enthusiastic narrator and with SLS ($t(107) = 2.454$, $p = 0.016$, $p_{\text{bonferroni}} = 0.094$, mean difference = 0.898, $d = 0.653$, 95% CI [0.118–1.188]), the group with a calm narrator and without SLS ($t(107) = 2.701$, $p = 0.008$, $p_{\text{bonferroni}} = 0.048$, mean difference = 1.001, $d = 0.728$, 95% CI [0.185–1.271]), and the group with a calm narrator and with SLS ($t(107) = 2.633$, $p = 0.010$, $p_{\text{bonferroni}} = 0.058$, mean difference = 1.002, $d = 0.744729$, 95% CI [0.171–1.286]).

Finally, in the case of transfer, there were three significant differences when looking at singular post-hoc tests, specifically between those who watched the videos with an enthusiastic narrator with SLS and those who watched the videos with a calm narrator, either without ($t(107) = 3.251$, $p = 0.002$, $p_{\text{bonferroni}} = 0.013$, mean difference = 1.935, $d =$

-1.260, 95% CI [-2.086—0.435]) or with SLS ($t(107) = 3.282, p = 0.002, p_{\text{bonferroni}} = 0.012$, mean difference = 1.996, $d = 1.301$, 95% CI [0.456–2.146]). There was also a difference between groups who watched the enthusiastic videos without SLS and those who watched the same type of videos but with SLS ($t(107) = -2.340, p = 0.024, p_{\text{bonferroni}} = 0.143$, mean difference = -1.471, $d = -0.959$, 95% CI [-1.809—0.109]), but the result did not remain significant after using a correction.

Higher proficiency group

In the higher English proficiency group, five interactions emerged as (marginally) significant, specifically participants' negative activation, paying attention to the video, intrinsic cognitive load, germane cognitive load, and level of certainty in one's correct answers on the immediate test (Appendix 17).

There was a marginally significant difference in negative activation when comparing the group without SLS and with an enthusiastic narrator with those who viewed videos with an enthusiastic narrator and with SLS ($t(99) = -1.689, p = 0.094$, mean difference = -0.363, $d = -0.480$, 95% CI [-1.049–0.088]) and with those who viewed videos with a calm narrator and without SLS ($t(99) = -2.123, p = 0.036$, mean difference = -0.455, $d = -0.602$, 95% CI [-1.172—0.033]). There was also a significant difference between participants who viewed clips with a calm narrator without and with SLS ($t(99) = 2.082, p = 0.040$, mean difference = 0.428, $d = 0.567$, 95% CI [0.021–1.113]).

In the calm narrator group, there was also a difference between those who did not and did have SLS ($t(98) = -2.190, p = 0.031, p_{\text{bonferroni}} = 0.185$, mean difference = -0.810, $d = -0.601$, 95% CI [-1.153—0.050]).

Regarding intrinsic cognitive load, two marginally significant differences were apparent—between the group watching the enthusiastic videos without SLS and those with calm videos with SLS ($t(98) = -1.891, p = 0.062, p_{\text{bonferroni}} = 0.370$, mean difference = -0.664, $d = -0.533$, 95% CI [-1.097–0.032]) and between those who watch the calm videos without and with SLS ($t(98) = -1.799, p = 0.075, p_{\text{bonferroni}} = 0.450$, mean difference = -0.616, $d = -0.494$, 95% CI [-1.043–0.055]).

Furthermore, there were four marginally significant differences in germane cognitive load. First, between those who watched calm videos without or with SLS ($t(98) = -2.431, p = 0.017, p_{\text{bonferroni}} = 0.101$, mean difference = -0.641, $d = -0.667$, 95% CI [-1.220—0.114]), then between participants who watched enthusiastic videos with SLS and those who watched calm videos with SLS ($t(98) = -1.833, p = 0.070, p_{\text{bonferroni}} = 0.419$, mean

difference = -0.496 , $d = -0.516$, 95% $CI [-1.080-0.047]$). Those who watched the videos with the enthusiastic narrator and without SLS also differed from participants who watched enthusiastic videos with SLS ($t(98) = 1.988$, $p = 0.050$, $p_{\text{bonferroni}} = 0.297$, mean difference = 0.544 , $d = 0.566$, 95% $CI [-0.005-1.136]$) and participants who watched videos with calm narrators and without SLS ($t(98) = 2.493$, $p = 0.014$, $p_{\text{bonferroni}} = 0.086$, mean difference = 0.689 , $d = 0.717$, 95% $CI [0.137-1.296]$). The last comparison was the only one in this whole subchapter that remained marginally significant even after applying the Bonferroni correction.

Lastly, there was a marginally significant difference in the level of certainty in correct answers between those who viewed calm videos without or with SLS ($t(97) = -1.688$, $p = 0.095$, $p_{\text{bonferroni}} = 0.567$, mean difference = -9.189 , $d = -0.467$, 95% $CI [-1.020-0.086]$).

3.3.3.5 Limitations and implications

In summary, the goals of Experiment 1 were to investigate the impact of the emotional stance of a disembodied narrator conveyed through voice only on learners who are watching educational videos in their non-native language, to explore the effects of same-language subtitles on learning from these videos, and to examine how results may vary based on participants' English proficiency. Our results show that while the narrator's emotional stance expressed through voice only significantly influenced participants' perceptions of the instructor, it did not impact their emotional state, interest, motivation, cognitive load, or overall learning performance when looking at the whole sample. However, the enthusiastic tone benefited learners with lower English proficiency but hindered those with higher proficiency, indicating a nuanced impact based on individual learner characteristics. Similarly, same-language subtitles (SLS) may slightly reduce extraneous cognitive load and improve transfer outcomes a week later among learners with lower English proficiency. These findings suggest that the effects of narrator emotional tone and SLS are not uniform across all learners, warranting further research to explore these nuances and potential boundary conditions.

There are several explanations for our results. While the calm and enthusiastic narrator groups perceived their narrator to be more calm or more enthusiastic compared to the other narrator, the group with the enthusiastic narrator rated the narrator higher on the calm scale than the enthusiasm scale. This finding implies that, although participants in the enthusiastic group recognized the narrator's enthusiasm more than those in the calm group, they ultimately perceived the enthusiastic narrator as more calm than enthusiastic, which could explain why there were no significant differences in emotional

state, interest, motivation, and cognitive load between the groups. If the enthusiastic narrator was not truly perceived as enthusiastic, the intended emotional impact might not have been achieved. The emotional tone, intended to energize and engage, may have been subdued by the perception of calmness, thereby diluting its effect. Further studies should incorporate their own narrations and interpretations of emotions through voice, as the participants' perception and subsequently our results might have been influenced by the specific performance of our chosen actor.

Another explanation is that voice alone may provide only minimal social and emotional cues, unlike facial expressions, gestures, and body language. This likely results in a small effect size, indicating that the impact on learning outcomes is limited. Consequently, larger sample sizes would be needed to detect any significant effects.

One limitation of our study is that the video content was not part of an academic course, resulting in lower participant motivation and potentially affecting performance. If the material had been more relevant to their coursework or interests, participants might have put more effort into learning, potentially leading to different results. While sustainable construction is an important topic for our future, it might not have been the most relevant to students from study programs unrelated to the topic. Most of our sample consisted of students from social science studies, whereas the topic of wood as a construction material would be more interesting to students from programs such as architecture, construction, and natural sciences. In the future, it would be beneficial to tailor the content to the participants' fields of study to enhance their engagement and effort, potentially yielding more accurate insights into the effects of our interventions.

Additionally, it is uncertain whether participants truly focused on the screen and the added text, which is particularly important for the SLS portion of the experiment and might explain the limited results. Future studies could benefit from using psychophysiological tools such as eye tracking and electrodermal activity measurements to ensure the interventions are effective and to gain further insights into students' emotional and cognitive processing.

Overall, the limited and null findings of the experiment prevent us from making causal conclusions. However, our experiment contributes significantly to the existing literature due to its robust methodology, including the use of objective rather than subjective measures of language proficiency and learning, the use of longer videos, and the assessment of knowledge both immediately after learning and one week later. Future research should continue to build on these methodological strengths while exploring the specific conditions under which SLS and narrator emotion might affect students'

emotional state, cognitive processing, and learning outcomes in a foreign language. This can be achieved through varied methodologies and diverse materials, including different content and actors displaying various emotions.

3.4 Pre-studies 2: Music rating

Before conducting the second main experiment, two pre-studies were made to choose and validate the two songs that will be added to the learning videos as the independent variable in Study 2.

3.4.1 Research hypotheses

There are no hypotheses for the first pre-study, but for the second one, three preliminary hypotheses were made:

Preliminary Hypothesis 4: The song that is to be used in the lively background music condition significantly elevates the participants' activation levels but not valence ratings compared to baseline.

Preliminary Hypothesis 5: The song that is to be used in the calm background music condition significantly lowers the participants' activation levels but not valence ratings compared to baseline.

Preliminary Hypothesis 6: The lively and calm song lead to significantly different activation levels but not valence ratings in participants.

3.4.2 Methodology

3.4.2.1 Research design

Both pre-studies were conducted as an online experiment with a within-subjects design in which participants listened to and rated the presented songs.

In the first pre-study, participants rated the emotional tone and energy level of 20 music tracks. The two songs that had the highest and lowest energy levels but similar ratings of emotional tone were selected and used in the next pre-study with the goal of determining whether the two songs affect the activation level of participants. In the second pre-study, therefore, participants rated how the two selected songs made them feel in both terms of valence and activation level on a Likert-type scale.

3.4.2.2 Participants

Demographic information from the samples of both pre-studies can be seen in Table 69. In the first pre-study, a convenience sample of 43 respondents aged from 18 to 60 ($M = 31.21$, $SD = 9.93$) participated, with 31 identifying as female and 12 as male with most being either a college student or employed, and no one having any reported difficulties of hearing. 14 participants were from Slovenia and the United Kingdom, three from Germany, two from the Check Republic, India, and the USA, and one each from France, Hungary, Lithuania, New Zealand, Portugal, and Singapore. A third of participants had formal musical training with an average of 7.93 years ($SD = 3.77$).

On the other hand, in the second pre-study, there were 66 participants with an average age of 32.76 ($SD = 11.13$), 42 being female and 24 being male. Most had at least a bachelor's degree and were either employed or self-employed. Almost half of the sample (31) was from Slovenia, eight from the UK, 4 from the USA, three from Germany and India, two from Italy, and one from Armenia, Australia, Bosnia and Herzegovina, Croatia, Ghana, Greece, Hong Kong, Hungary, Ireland, Jordan, Mexico, Montenegro, Norway, Poland, and Uganda. One person reported they have trouble hearing in noisy places and 30% of the sample had musical training ($M = 8.20$, $SD = 3.14$).

Table 69: Demographics from both pre-studies

	First pre-study ($N = 43$)	Second pre-study ($N = 66$)
Gender		
Female	31 (72.09%)	42 (63.64%)
Male	12 (27.91%)	24 (36.36%)
Education		
Primary education	1 (2.33%)	
Secondary education	2 (4.65%)	7 (10.61%)
Bachelor's degree	15 (34.88%)	19 (28.79%)
Master's degree	19 (44.19%)	27 (49.91%)
Doctorate degree	6 (13.95%)	13 (19.70%)
Status		
High school student	1 (2.33%)	
University student	18 (41.86%)	13 (19.70%)
Employed / self-employed	22 (51.16%)	49 (74.24%)
Unemployed		1 (1.52%)

Other	1 (2.33%)	1 (1.52%)
Undisclosed	1 (2.33%)	2 (3.03%)
Formal musical training		
Yes	15 (34.88%)	20 (30.30%)
No	28 (65.23%)	46 (69.70%)
Hearing difficulties		
Yes	0 (0.00%)	1 (1.52%)
No	43 (100.00%)	64 (96.97%)

Note. *n* (%)

3.4.2.3 Material

The candidate chose 20 songs with different levels of activation – 10 calm and 10 lively (see Appendix 4 for details) – from the open-source audio library pixabay.com, making sure all music tracks were copyright free and available for noncommercial use. The goal of the selection process was to find songs that are representative of the type of songs that are usually present as background music in learning videos, which are predominantly positive, electronic, and instrumental. While this type of background music, which is defined as music that accompanies the dialogue or action of a motion picture, radio or television drama (Merriam-Webster, n.d. - a), does not have a specific name, it is similar but not the same as ambiental or incidental music, which is music intended to create a certain atmosphere or project a particular mood in the audience of a play or video production, such as helping them feel more energized or relaxed (Merriam-Webster, n.d. - b). Common keywords used to find all songs were: “background music, music for videos, music for Youtube videos, vlog music, podcast music, ambiental, corporate, acoustic, instrumental”. In searching for lively songs, we used additional keywords such as “upbeat, energetic, powerful, uplifting, bright, happy, electronic, fast, very fast”, while for calm songs the following additional keywords were used: “relaxing, calm, peaceful, slow, medium slow, medium tempo.” Songs were selected based on the following criteria: 1) instrumental music (no lyrics or human vocalisations present), 2) likely to be unfamiliar to participants, 3) written in major mode as these types of musical excerpts are usually perceived as happy (Gagnon and Peretz, 2003), 4) stable mood and activation levels for most of the track’s duration, 5) duration of 2 minutes and more, and 6) modern. Tempo and mode were analysed with an online song key and BPM finder tool tunebat.com.

For the pre-studies, songs were presented in a randomized order and assigned a letter (e.g., Song A, Song B, ... Song U) so that the title of the songs would not affect the respondent's perception of the music track. In the first pre-study, participants were instructed to listen to each song for at least 30 seconds (but more was recommended) and to answer the related questions as soon as they stopped listening to each song.

In the second pre-study, only two musical tracks were used based on the ratings of the first pre-study. In this case, respondents needed to listen to the songs for 90 seconds or more. The order in which the songs were presented was randomized.

3.4.2.4 Instruments

After each song in the first pre-study, two rating scales were shown – one asking participants to rate the energy level of the song on a 9-point scale (1 – Extremely calm, 5 – Somewhere in between, 9 – Extremely lively) and one asking them to rate the emotional tone or mood of the corresponding song, also on a 9-point scale (1 – Extremely negative, 5 – Somewhere in between, 9 – Extremely positive). Both questions had a prompt reminding participants to respond based on the presented songs and not how the songs made them feel, together with an explanation. The first question had the following explanation: *For example, a song with a low energy level brings to mind words such as "calm", "relaxing" or "boring", while a song with a high energy level brings to mind words such as "lively", "tense" or "exciting;"* while the question regarding the songs' mood had this explanation: *For example, a song with a negative tone brings to mind words such as "sad", "angry" or "boring", while a song with a positive tone brings to mind words such as "happy", "exciting" or "calming".* At the end of the survey, respondents were presented with demographic questions about their gender, age, education, student/employment status, country of residence, years of formal musical training, and presence of any hearing difficulties.

The second pre-study was similar to the first one but included only two selected songs and the respondents rated how they felt and not what they thought that the music expressed or conveyed. Participants were asked to rate how pleasantly do they feel at the moment (valence) and what is their current level of activation regardless of whether the feeling is pleasant or unpleasant (activation level) three times: before listening to the songs and after listening to each song for at least 90 seconds. Both questions had a 9-point rating scale (1 – Extremely unpleasant/low activation, 5 – Somewhere in between, 9 – Extremely pleasant/high activation) and additional examples to help respondents understand the questions. The valence question had the following explanation:

Examples of unpleasant feelings are nervousness, frustration, boredom, or sadness, while examples of pleasant feelings are enthusiasm, joy, contentment, or relaxation, while the following examples were added to the second question: Examples of low activation are relaxation, boredom, contentment, or sadness, and examples of higher activation are alertness, enthusiasm, nervousness, or frustration. The survey ended with demographical questions.

3.4.2.5 Data collection

People over the age of 15 were invited to participate in the candidate's and InnoRenew CoE's social media. Both surveys were made in Slovene and English language, conducted online, and displayed on the platform 1ka.si (Faculty of Social Sciences, University of Ljubljana, 2022).

Participants were able to participate in one or both pre-studies. The first pre-study lasted for 10 to 20 minutes and was conducted in the first half of December 2022 while the second pre-study was 5 minutes long and was conducted in the second half of the same month. Respondents received no incentives for participation in either pre-study.

3.4.2.6 Data analysis

Data was analysed using the open-source software jamovi (The jamovi project, 2022). Descriptive statistics were made for ratings from both pre-studies, together with paired samples *t*-tests.

3.4.3 Results and interpretation

The results of the first pre-study for Study 2 are presented in Table 70 which shows the perceived energy level and emotional tone of each of the 20 songs ranked from lowest to the highest energy level rating. Ideally, the two selected songs would be very different in energy levels but have no difference in the emotional tone rating, which was not possible in this case as the songs with lower energy level ratings had also lower emotional tone ratings, although not as significantly. For the next step, songs O and F were chosen as they had the second lowest and highest energy level ratings respectively, but the difference in their emotional tone ratings was smaller compared to the songs with the lowest/highest energy level ratings. Song F had significantly higher ratings of both energy level ($t(42) = 24.48$, $p < .001$, mean difference = 5.42, 95% CI [4.97–5.87], $d = 3.73$, 95% CI [2.88–4.58]) and emotional tone compared to Song O

($t(42) = 4.90$, $p < .001$, mean difference = 1.79, 95% CI [1.05–2.53], $d = 0.75$, 95% CI [0.41–1.08]).

Table 70: Songs with their corresponding energy level and emotional tone ratings

Song	Calm / lively	Energy level rating		Emotional tone rating	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
H	Calm	2.09	0.95	4.33	1.71
O	Calm	2.12	1.22	5.21	1.79
E	Calm	2.30	1.12	4.77	1.81
G	Calm	2.30	1.24	4.93	1.67
L	Calm	2.47	1.18	4.40	1.83
T	Calm	2.98	1.08	5.44	1.68
D	Calm	3.07	1.52	5.16	1.57
I	Calm	3.37	1.63	5.93	1.39
U	Calm	4.21	1.28	5.37	1.45
C	Calm	4.33	1.39	5.56	1.22
P	Lively	6.88	0.98	6.37	1.27
A	Lively	7.05	1.02	6.84	1.54
M	Lively	7.12	1.12	6.47	1.50
N	Lively	7.21	1.19	6.74	1.43
J	Lively	7.23	0.95	6.65	1.72
S	Lively	7.33	1.17	7.07	1.14
B	Lively	7.49	1.26	7.09	1.41
R	Lively	7.49	0.96	7.12	1.40
F	Lively	7.53	0.96	7.00	1.31
K	Lively	7.60	1.31	7.23	1.54

Note. Songs in bold were selected to use in the main experiment.

Table 70 contains results from the second pre-study of Study 2, indicating that after listening to Song F, participants felt significantly more activated ($M = 6.03$, $SD = 1.47$) compared to baseline ($M = 4.97$, $SD = 1.70$) and after listening to Song O ($M = 4.41$, $SD = 1.61$) they felt significantly less activated compared to baseline. Regarding valence, there were no significant differences after either Song F ($M = 6.35$, $SD = 1.42$) or Song O ($M = 6.05$, $SD = 1.41$) compared to baseline ($M = 5.95$, $SD = 1.60$). These results confirm Preliminary Hypotheses 4 and 5.

To further verify whether there are significant differences between the two songs, we transformed raw ratings into change scores by subtracting baseline ratings from ratings of Song F and Song O (e.g., Song F activation level – baseline activation level, Song O valence – baseline valence). As Table 71 shows, there was a significant difference in change scores of activation level ratings between Song F ($M = 1.06$, $SD = 1.78$) and Song O ($M = -0.56$, $SD = 1.85$), while there was no significant difference in valence change scores of Song F ($M = 0.39$, $SD = 1.82$) and Song O ($M = 0.09$, $SD = 1.80$), confirming Preliminary Hypothesis 6.

Based on the results of this validation pre-study, we can deduct that the two selected songs caused significant changes in participants' activation levels but not in feelings of pleasantness, meaning that they can be used in Study 2 as two variations of the independent variable; Song F for the condition with the lively background music and Song O for the condition with the calm background music.

Table 71: Paired t -tests comparing valence and activation level ratings of Song F and O with baseline ratings and change scores of Song F and O between themselves

Variables	t	p	Mean difference [95% CI]	d [95% CI]
$Act_F - Act_{baseline}$	4.84	< .001	1.06 [0.62–1.50]	0.60 [0.33–0.86]
$Val_F - Val_{baseline}$	1.76	0.084	0.39 [–0.05–0.84]	0.22 [–0.03–0.46]
$Act_O - Act_{baseline}$	–2.46	0.016	–0.56 [–1.02– –0.11]	–0.30 [–0.55– –0.06]
$Val_O - Val_{baseline}$	0.41	0.684	0.09 [–0.35– 0.53]	0.50 [–0.19–0.29]
$Act_{F-cs} - Act_{O-cs}$	6.89	< .001	1.62 [1.15–2.09]	0.85 [0.56–1.13]
$Val_{F-cs} - Val_{O-cs}$	1.33	0.189	0.30 [–0.15–0.76]	0.16 [–0.08–0.41]

Note. $df = 65$, cs – change score

3.5 Study 2: Experiment on the effect of background music

3.5.1 Research hypotheses

Three hypotheses were made for Study 2:

H9: Participants learning with lively music will have higher levels of positive activating emotions than participants learning without music or with calm music.

H10: There will be significant differences in cognitive load between participants learning without music and participants learning with lively and calm music.

H11: There will be significant differences in learning outcomes between participants learning without music and participants learning with lively and calm music.

3.5.2 Methodology

3.5.2.1 Research design

Similarly to Study 1, Study 2 was carried out as a quantitative experimental research utilizing both descriptive and causal experimental methods and a between-subjects design. However, Study 2 focused only on one factor and included a control group together with two experimental groups. Participants in the control group watched videos with no background music added, while the participants in the two experimental groups learned from videos that had added background music – either calm or lively, exciting background music. The allocation of participants to each group was done through a randomized process. All participants underwent an identical experimental procedure, with the sole difference being the nature of the videos they viewed. Following a week after the initial experiment, participants were invited to participate in the second phase of the study, involving responding to the questions of the same knowledge test they has taken seven days prior.

3.5.2.2 Participants

A convenience sample of 307 students took part in Study 2, with 299 (97.39%) of them who responded the experiment in full and 8 (2.61%) who did not finish the experiment for varying reasons, providing only partial data. 102 (33.22%) participants were part of the control group, 105 (34.20%) were part of the group who viewed the videos with added calm music, and 100 (32.57%) learned from videos with lively background music. Most participants came from various faculties and 40 programs from

the University of Primorska (UP; 175 or 57.00%) and the University of Ljubljana (UL; 92 or 29.97%) from Slovenia, and Oregon State University from the United States of America (OSU; 29 or 9.45%), but for 11 (3.58%) students, their university is unknown. Students from the University of Primorska were mostly studying in social sciences programs, while students from the University of Ljubljana and Oregon State University were most commonly part of life sciences study programs. Three-quarters of the participants were students at the Bachelor's level (231 or 75.24%), 41 were Master's students (13.36%), 22 students at Doctoral level (7.17%), six noted that they were on a break or their absolvent stage (1.95%), and for seven participants there is no available data.

Participants' demographics are presented in more detail in Table 72, providing data of the sample as a whole and divided by universities. However, information from seven participants with partial data is not included.

Table 72: Demographics divided by university and in total

	UP (N = 175) <i>n</i> (f%)	UL (N = 92) <i>n</i> (f%)	OSU (N = 29) <i>n</i> (f%)	Undisclosed (N = 6) <i>n</i> (f%)	Total (N = 302) <i>n</i> (f%)
Gender					
Female	131 (43.38%)	53 (17.55%)	11 (3.64%)	5 (1.66%)	200 (66.23%)
Male	40 (13.25%)	39 (12.91%)	17 (5.63%)	1 (0.33%)	97 (32.12%)
Non-binary	1 (0.33%)		1 (0.33%)		2 (0.66%)
Undisclosed	3 (0.99%)				3 (0.99%)
Study fields (KLASIUS-P-16)					
Education	63 (21.00%)	3 (1.00%)			66 (22.00%)
Arts and Humanities	6 (2.00%)				6 (2.00%)
Social sciences, journalism, and information	22 (7.33%)				22 (7.33%)

Business, administration, and law	51 (17.00%)	2 (0.67%)		1 (0.33%)	54 (18.00%)
Natural sciences, mathematics, and statistic	1 (0.33%)	10 (3.33%)	6 (2.00%)		17 (5.67%)
Information and Communication Technologies (ICTs)	8 (2.67%)				8 (2.67%)
Engineering, manufacturing and construction	7 (2.33%)	76 (25.33%)	18 (6.00%)	2 (0.67%)	103 (34.33%)
Agriculture, forestry, fisheries, and veterinary			3 (1.00%)		3 (1.00%)
Health and welfare	17 (5.67%)	1 (0.33%)		1 (0.33%)	19 (6.33%)
Services			2 (0.67%)		2 (0.67%)
Study level					
Bachelor's	138 (45.70%)	75 (24.83%)	18 (5.96%)	1 (0.33%)	232 (76.82%)
Master's	23 (7.62%)	14 (4.64%)	4 (1.32%)	1 (0.33%)	42 (13.58%)
Doctorate	13 (4.30%)		7 (2.32%)	2 (0.66%)	22 (6.95%)
Absolvent	2 (0.66%)	4 (1.32%)			6 (1.99%)

Given the increased number of study programs to which participating students belonged, Table 72 categorizes them using the KLASIUS-P-16 classification (Statistical Office of the Republic of Slovenia, 2023b). KLASIUS-P-16 is the official classification in Slovenia that groups different study programs into classification groups or categories regarding the similarity of subject-specific characteristics of their content (Statistical Office of the Republic of Slovenia, 2023a). Among all the classification categories, the Engineering, manufacturing, and construction group stands out, as students within this category possess greater familiarity with the subject matter of the instructional videos.

This group comprises individuals from study programs like Architecture, Civil Engineering, Wood Science, and Wood Engineering, indicating a potentially higher level of expertise on the topic. The full list of study programs of the participants can be seen in Appendix 18.

The majority of participants came from Slovenia (243 or 79.15%), followed by the United States of America (21 or 6.84%), North Macedonia (15 or 4.89%), and Serbia (6 or 1.95%), with two participants or 0.65% being from Bosnia and Hercegovina, Canada, and Germany, and one participant from each of the following countries: Argentina, Belarus, Croatia, Ghana, Hungary, Nigeria, Norway, Peru, Poland, Russia, South Korea, Spain, Thailand, Tunisia, and Ukraine. For one of the participants, their country of origin is unknown. 241 or 78.50% of the participants participated in the experiment in the Slovene language, while the rest (66 or 21.50%) had the instruments presented in the English language. However, based on the information on country of origin, we can deduce that fewer than seven percent of participants were individuals whose native language is English. This proportion is insufficient for a meaningful comparison of outcomes between those who are native English speakers and those who are not.

On average, the participants were about 22.64 years old ($Mdn = 21$, $SD = 4.97$), and the ages ranged from 18 to 56. Split between the three universities, the average age of participants from UP was 22.46 ($Mdn = 21$, $SD = 5.25$), the average age of students from UL was 21.71 ($Mdn = 21$, $SD = 3.16$), while the students from OSU were older than their Slovenian counterparts with an average age of 26.38 ($Mdn = 23$, $SD = 6.47$). Among the participants, two-thirds (200 or 66.23%) identified as women, 97 (32.12%) as men, two (0.65%) noted they were non-binary, and nine (2.93%) chose to not disclose their gender or stopped the experiment before providing this information.

296 or 98.67% of participants who answered demographical questions (300) reported not having any difficulties in hearing, while four participants mentioned some kind of problems, namely reduced hearing or deafness in one ear, not hearing well at times, and sound sensitivity (hyperacusis). When asked about formal music training, 213 individuals, constituting 71.00% of the respondents, indicated that they did not possess any formal musical training. Meanwhile, 47 respondents, accounting for 15.57% of the total, reported having received formal musical training for a duration ranging from one to six years (with six years representing the length of lower music school in Slovenia). Additionally, 22 respondents, or 7.33% of the total, disclosed having undergone seven to eight years of formal musical training (eight years corresponding to the duration of higher music school in Slovenia). Furthermore, 17 respondents, constituting 5.67% of

the sample, reported a range of nine to fourteen years of musical training. Lastly, one respondent had 30 years of musical training.

Regarding the delayed testing, a total of 118 students (38.44% of all students who participated in the main experiment), took part in the second phase of the study, with 72 students being from UP, 34 from UL, 9 from OSU, and three from an undisclosed university. This included 40 students who watched the videos without added background music, 35 students who viewed videos with calm music, and 43 students from the lively music group.

3.5.2.3 Material

The study utilized a series of custom-made learning videos on wood as a building material in English, incorporating slides and narration, and in some cases, background music. The content and visuals were the same as in the first experiment, meaning that there were five different videos with a total duration of 24 minutes. This segmented approach allowed participants to take brief breaks as needed, maintaining focus on the material, and thus enhancing participant engagement and preventing attention loss. Moreover, the use of multiple videos enabled the assessment of participants' emotional and mental states at various points.

There were three versions of the videos, resulting in a total of fifteen unique videos – learning videos without added background music and two videos with added music, one a calm song and one a lively song. The videos in the no music condition were the same as the videos used in the first experiment for the enthusiastic and no subtitles condition. The videos were made using Microsoft PowerPoint and the narration was added with the Audacity® program. The background music to be added in the calm and lively music conditions was selected and validated in the previously described Pre-studies 2. For the calm music condition, the song titled “Hopeful Slow Atmospheric Meditation” by composer Ashot-Danielyan-Composer was added. It is a song that is 3 minutes and 55 seconds long and composed in E major with a tempo of 90 beats per minute. For the lively music condition, the song “Fun and Happy” by the composer Alex MakeMusic was chosen. This song lasts for 2 minutes and 8 seconds, has a tempo of 125 beats per minute, and is composed in B♭ major. Both songs were instrumental in nature and did not include any vocals, and the main difference was in their rhythm, one being calm, and the other lively and upbeat. The same song was integrated into all five videos within the video series, based on the experimental condition. The music's conclusion was managed such that it gradually faded out, followed by a 5-second fade-

in before restarting from the beginning. The same videos as in the no music condition were used in all conditions, but the songs were added with the Windows Media Player software. Again, the use of prevalent and accessible software was intentional to ensure reproducibility by educators and content creators possessing only basic computer programs.

As in the first study, the learning material included static representational pictures or graphics, minimal text, and narration made by a female with a Standard American English accent (the enthusiastic version). No subtitles or videos of the instructor were present. The first video, lasting 3 minutes and 13 seconds, introduced participants to wood as a material and the concept of service life. The second, almost 6-minutes-long video, focused on wood degradation processes and how to combat them. The last three videos covered topics, such as material properties and selection (6:38 minutes long), protective design measures (3.49 minutes long), and wood maintenance (4:30 minutes long).

3.5.2.4 Instruments

The majority of the instruments were the same as those used in Study 1, but some instruments were omitted and some were added in the case of this study, so all instruments will be described again. There were also some slight changes in the order of different instruments and all materials were translated in Slovene and in English, but not in Norwegian. Again, the survey was mostly made up of questionnaires that were previously validated in international studies, together with the same knowledge test that was used and validated in the first experiment. As was the case in Study 1, the reliability of each instrument will be reported using McDonald's ω , as it is superior to Cronbach's α when the assumption of tau equivalence is not met and the same when it is (McNeish, 2017).

The survey assessed these factors: pre-existing subject knowledge, personal perception of pre-existing knowledge and experience level, previous interest in the subject, emotional state, mental exertion, self-evaluated learning, video perception, self-evaluated effect of music, cognitive load, motivation and interest in the topic, personality, knowledge evaluation, demographic attributes, subjective and objective English language proficiency, and some questions regarding studying with music.

Knowledge (pre-test and post-test): These tests were the same as used in Study 1. A pre-test with eight multiple-choice questions (Appendix 5) was first administered to test participants' preexisting knowledge of the subject covered by the videos. These

questions were different than the ones used for outcome testing to avoid priming students to the specific content of the questions. On top of four possible answers, the pre-test questions included an “I do not know” option so the participants did not have to guess. One point was assigned only for a correct answer. No feedback was given to the participants regarding their answers or scores.

Table 73 displays the difficulty indexes, which represent the proportion of participants who answered each item correctly, of the pre-test questions in Study 2. The results are very similar to those from the previous experiment, mainly because indexes are quite low. This is not problematic as it is a pre-test before the intervention. The reliability of the pre-test is higher than it was in Study 1, especially for the English version ($\omega = 0.591$; $\omega_{Slo} = 0.498$; $\omega_{Eng} = 0.763$).

Table 73: Item difficulty indexes of pre-test questions in Study 2

Question	IDI
PT1	0.42
PT2	0.28
PT3	0.26
PT4	0.08
PT5	0.22
PT6	0.30
PT7	0.18
PT8	0.37

The post-test consisted of 29 multiple-choice questions (Appendix 6) with four alternative answers and no “I do not know” option. The whole knowledge test includes 19 retention and 10 transfer questions. In the Slovene version of the survey, questions and answers were written in both Slovene and English due to the presence of some technical terminology in the videos that would be hard to know in Slovene if someone is unfamiliar with the subject matter. One point per correct answer was assigned, with the maximum score thus being 29 points overall (19 for the retention test and 10 for the transfer test). The questions’ order was consistent in both the main and delayed part of the experiment and in neither testing session participants received no feedback on their test performance.

Similarly to Study 1, Study 2 also included an extra question after each post-test question, prompting participants to express their confidence level in their response as a

percentage. To standardize the understanding of percentage values among all participants, the question also had an explanation that due to there being four possible answers, a response of 25% would reflect a complete guess on their part. The same was repeated in the delayed part of the experiment.

Table 74 presents the item difficulty indexes for both the initial and delayed parts of Study 2, alongside participants' confidence levels when answering correctly. Similar to the outcomes in Study 1, item difficulty indexes spanned from 0.25 to 0.82 in the first phase and from 0.24 to 0.85 in the delayed phase, again verifying that the knowledge test was appropriate. The overall difficulty indexes in both parts of the experiment were approaching the optimal index of 0.60 ($M_{IDI_{main}} = 0.55$, $SD_{IDI_{main}} = 0.15$; $M_{IDI_{delayed}} = 0.55$, $SD_{IDI_{delayed}} = 0.17$) (Bucik, 1997). Equivalent results were obtained when analysing the retention ($M_{IDI_{main}} = 0.55$, $SD_{IDI_{main}} = 0.14$; $M_{IDI_{delayed}} = 0.54$, $SD_{IDI_{delayed}} = 0.14$) and transfer ($M_{IDI_{main}} = 0.55$, $SD_{IDI_{main}} = 0.19$; $M_{IDI_{delayed}} = 0.56$, $SD_{IDI_{delayed}} = 0.21$) sections of the test individually.

Table 74: Item difficulty indexes and confidence levels of correct responses on post-test questions

Question	Type of knowledge	Study 2 – part 1 (N = 224)		Study 2 – part 2 (N = 94)	
		IDI	Confidence	IDI	Confidence
R1	Retention	0.73	85.44%	0.72	79.13%
R2	Retention	0.82	78.01%	0.84	71.42%
R3	Retention	0.28	60.71%	0.28	61.91%
R4	Retention	0.51	62.33%	0.50	70.00%
R5	Retention	0.54	71.97%	0.69	69.00%
R6	Retention	0.47	67.28%	0.43	66.80%
R7	Retention	0.52	72.11%	0.51	70.78%
R8	Retention	0.61	63.82%	0.57	56.00%
R9	Retention	0.41	71.21%	0.47	70.95%
R10	Retention	0.56	71.26%	0.51	69.60%
R11	Retention	0.50	64.80%	0.50	67.15%
R12	Retention	0.50	87.55%	0.47	88.20%
R13	Retention	0.77	75.53%	0.72	68.92%
R14	Retention	0.46	61.89%	0.48	60.63%
R15	Retention	0.55	61.47%	0.58	63.00%
R16	Retention	0.57	71.46%	0.43	64.41%

R17	Retention	0.74	68.53%	0.71	66.86%
R18	Retention	0.49	71.91%	0.46	68.76%
R19	Retention	0.36	71.53%	0.36	76.05%
T1	Transfer	0.67	79.34%	0.64	74.97%
T2	Transfer	0.62	66.84%	0.72	69.38%
T3	Transfer	0.26	68.39%	0.24	59.11%
T4	Transfer	0.65	70.28%	0.75	66.32%
T5	Transfer	0.52	56.07%	0.52	56.69%
T6	Transfer	0.76	78.31%	0.77	79.63%
T7	Transfer	0.51	65.11%	0.48	68.81%
T8	Transfer	0.81	73.91%	0.85	76.35%
T9	Transfer	0.49	70.87%	0.37	73.95%
T10	Transfer	0.25	66.29%	0.30	64.14%

Note. IDI – item difficulty index

During the main part of the experiment, the knowledge test demonstrated satisfactory internal consistency both when looking at the whole sample and when dividing the sample based on survey language ($\omega = 0.782$; $\omega_{Slo} = 0.763$; $\omega_{Eng} = 0.822$) (McNeish, 2017). Upon isolating the retention and transfer tests, it can be deduced that the retention test by itself still has acceptable levels of internal consistency ($\omega = 0.725$; $\omega_{Slo1} = 0.700$; $\omega_{Eng1} = 0.773$), while for the transfer test reliability was notably lower ($\omega = 0.498$; $\omega_{Slo1} = 0.471$; $\omega_{Eng1} = 0.581$).

The same trend can be seen in the post-test a week after the initial phase. In the delayed part of the experiment, the knowledge test as a whole retained its satisfactory level of internal consistency ($\omega_2 = 0.745$; $\omega_{Slo2} = 0.728$; $\omega_{Eng2} = 0.772$). When looking at the retention and transfer assessments separately, the retention test exhibited still acceptable levels of internal consistency ($\omega_2 = 0.638$; $\omega_{Slo2} = 0.728$; $\omega_{Eng2} = 0.731$), which cannot be said for the transfer test ($\omega_2 = 0.436$; $\omega_{Slo2} = 0.408$; $\omega_{Eng2} = 0.386$).

As in the previous experiment, after finishing the test, participants rated their subjective test performance on a 7-point Likert-type scale (1 – Very poorly, 4 – Neutral, 7 – Very well). However, two additional learning-related variables were added to Study 2 to gain more knowledge on participants' self-evaluation of the learning experience. After watching the last video, students were asked to state how much of the content in the videos was new knowledge for them as a percentage of the total content. In addition, they also had to assess their perceived level of learning from the videos on a 7-point Likert-type scale (1 – Very poorly, 4 – Neutral, 7 – Very well).

Subjective pre-existing knowledge, experience, and interest: Similarly to Study 1, participants assessed their familiarity with the topic of wood as a construction material, their degree of experience in working with wood, and their level of interest in the subject, all using a 7-point scale (1 – Very low/I have never worked with wood/I am not interested at all, 4 – Moderate/I rarely work with wood/Neither interested nor not interested, 7 – Very high/I work with wood very often/Very interested). Participants were asked about their interest in the topic again in the second part of the experiment.

English language: The same instruments were used as in Study 1, with the distinction that the language section was positioned at the conclusion of the survey rather than at the start. This change was made to enable participants to view the educational videos sooner and with a clearer perspective. The language segment included three questions and a brief English assessment. First, participants were asked to specify the total count of languages they understand, including their native tongue. Then, they ranked their understanding of English relative to the other languages they were familiar with. To ensure all participants understood the question, they were provided with an example: “If you indicated above that you understand 4 languages and you think you understand English better than the other two foreign languages but less than your mother language, please indicate the number 2.” Lastly, the third question prompted participants to rate their capacity to understand spoken English on a scale from 1 (very low) to 7 (very high).

The standardized *Lexical Test for Advanced Learners of English* or *LexTALE* (Lemhöfer and Broersma, 2012) was used for an objective assessment of English proficiency as it has been demonstrated as a good indicator of overall English proficiency (as indicated by more comprehensive proficiency assessments like the TOEIC and the Quick Placement Test). In LexTALE, participants must determine whether the 60 presented words are actual English terms or not (20 words are not real words). It demonstrated a good internal consistency ($\omega = 0.881$; $\omega_{Slo} = 0.854$; $\omega_{Eng} = 0.917$ – two items in the English version had no variability so they were omitted from the analysis).

Emotional outcomes: The three instruments used to assess participants' emotional state were the same as those used in Study 1. The first two were single-item scales designed to assess the two dimensions of the circumplex model of *core affect* – specifically, the level of arousal/activation and the degree of pleasure/valence (Russell, 1980; Russell et al., 1989). Each of these scales was administered six times – just before the first video and then after each subsequent video viewing. A modified version of these scales was used as originally, the items measuring pleasure and arousal were structured

as a singular affect grid, where respondents indicated their current mood by marking a grid with columns representing the pleasure score and rows signifying the arousal score. In both our studies, however, we opted for two distinct single-item scales. The valence item displayed good internal consistency ($\omega = 0.913$; $\omega_{Slo} = 0.916$; $\omega_{Eng} = 0.906$) and was structured as follows: "How pleasantly do you feel at the moment? Examples of unpleasant feelings are nervousness, frustration, boredom, or sadness, while examples of pleasant feelings are enthusiasm, joy, contentment, or relaxation." The arousal item was presented in a similar way: "What is your level of activation at the moment, regardless of whether the feeling is pleasant or unpleasant? Examples of low activation are relaxation, boredom, contentment, or sadness, and examples of higher activation are alertness, enthusiasm, nervousness, or frustration." This item also exhibited good internal consistency ($\omega = 0.914$; $\omega_{Slo} = 0.920$; $\omega_{Eng} = 0.892$). Participants rated both items using a 9-point Likert-type scale (1 – Extremely unpleasant/low activation, 2 – Very unpleasant/low activation, 3 – Unpleasant/Low activation, 4 – Somewhat unpleasant/low activation, 5 – Somewhere in between, 6 – Somewhat pleasant/high activation, 7 – Pleasant/High activation, 8 – Very pleasant/high activation, 9 – Extremely pleasant/high activation).

The third instrument used for measuring the affective state of participants was the *Positive Activation, Negative Activation and Valence Short Scale* (PANAVA-KS; Schallberger, 2005), which is grounded in the dual activation systems model of affect (Watson and Tellegen, 1985) and comprises three dimensions: positive activation (four items; $\omega = 0.858$, $\omega_{Slo} = 0.863$, $\omega_{Eng} = 0.845$), negative activation (four items; $\omega = 0.872$, $\omega_{Slo} = 0.869$, $\omega_{Eng} = 0.886$), and valence (two items; $\omega = 0.788$, $\omega_{Slo} = 0.817$, $\omega_{Eng} = 0.684$). Participants were instructed to rate their current emotional state on a 7-point bipolar Likert-type scale ranging from -3 to $+3$ (e.g., "satisfied – dissatisfied"; "full of energy – no energy", "stressed – relaxed"). The PANAVA-KS was used twice during the experiment: immediately before viewing the first video (baseline measurement) and following the last video.

Cognitive outcomes: Cognitive outcomes of participants were also assessed using the same two instruments that were used in Study 1 – the single-item measure of *subjective mental effort* (Paas, 1992) and the *Cognitive Load Questionnaire* (Klepsch et al., 2017). The first instrument asks participants to rate the extent of mental effort they exerted in understanding the video content on a 9-point scale (1 – Very, very low mental effort, 2 – Very low mental effort, 3 – Low mental effort, 4 – Rather low mental effort, 5 – Neither low nor high mental effort, 6 – Rather high mental effort, 7 – High mental effort,

8 – Very high mental effort, 9 – Very, very high mental effort). Due to its brevity and ease of implementation, it was used after the viewing of each video (five times in total). The level of internal consistency was high ($\omega = 0.938$; $\omega_{Slo} = 0.941$; $\omega_{Eng} = 0.928$).

After watching all the videos, participants also reported their subjective cognitive load through the Cognitive Load Questionnaire, a questionnaire that distinguishes between distinct forms of cognitive load (Klepsch and Seufert, 2020). Specifically, it differentiates between intrinsic (two items; $\omega = 0.619$; $\omega_{Slo} = 0.624$; $\omega_{Eng} = 0.594$; e.g., “Learning from the videos was very complex”), extraneous (three items; $\omega = 0.790$; $\omega_{Slo} = 0.787$; $\omega_{Eng} = 0.801$; “The design of the learning videos was very inconvenient for learning”), and germane cognitive load (two items; $\omega = 0.525$; $\omega_{Slo} = 0.622$; $\omega_{Eng} = 0.510$; “I made an effort, not only to understand several details, but to understand the overall context.”). As was the case in Study 1, an item measuring germane cognitive load (“The learning task consisted of elements supporting my comprehension of the task.”) was omitted as it was not relevant to our experiment. Participants indicated their level of agreement with the statements using a 7-point Likert-type scale (1 – Strongly disagree, 2 – Disagree, 3 – Somewhat disagree, 4 – Somewhere in between, 5 – Somewhat agree, 6 – Agree, 7 – Strongly agree). The same scale was used also when participants rated their interests, motivation, video experience, and personality.

Situational interest: The survey included the same instrument as in Study 1 to understand how much the videos induced participants’ interest in the topic (Rotgans and Schmidt, 2011). The short tool consists of six questions, with one of them being reverse-scored, and demonstrated good reliability ($\omega = 0.884$; $\omega_{Slo} = 0.889$; $\omega_{Eng} = 0.875$). Participants expressed their agreement level with each statement using a 7-point scale.

Intrinsic Motivation: The questionnaire by Isen and Reeve (2005) was also kept to assess participants’ motivation for watching the videos. As in Study 1, the wording of the items was changed in a way to match the study’s context. The instrument displayed excellent internal consistency ($\omega = 0.884$; $\omega_{Slo} = 0.889$; $\omega_{Eng} = 0.875$). Participants indicated their agreement level with eight statements using a 7-point scale.

Video experience: The last set of questions that were kept from Study 1 were statements regarding the participants’ experience with videos taken from recent studies on the impact of emotional design on learning from multimedia (e.g., Lawson et al., 2021a, 2021c). The five statements do not represent a single factor but give some insight into the viewers’ enjoyment of the videos, motivation to pay attention to them, the difficulty of the presented information, the level of effort they put into understanding the provided content, and whether they would like to have more lessons that are similar to

the one they experienced. The same 7-point Likert-type scale was used for rating the level of agreement with the statements.

In addition, participants also rated how pleasant or unpleasant and passive or active they perceived the videos on a 7-point scale (1 – Very unpleasant/passive, 4 – Neutral, 7 – Very pleasant/active).

Perception of background music influence: As background music was only introduced in Study 2, a new question was added, inquiring about how the music during the videos influenced the participants. Seven response options were provided: a) "It was distracting me from the video content," b) "It helped me concentrate on the video content," c) "It relaxed me," d) "It energized me," e) "It had no influence on me," f) "Other," and g) "I did not notice any music during the videos." If the respondents chose the answer option "Other," they were prompted to provide more information. The open-ended responses were coded and categorized by the candidate.

Experiences with listening to music during learning: Additional questions were introduced to gather insights about participants' habitual use of music while studying. Initially, students were requested to indicate their frequency of studying with background music. The available choices were: "Never," "Rarely," "Sometimes," "Often," "Very often," and "Always." If respondents selected any option other than "Never," two supplementary questions emerged. The first question inquired about the primary reason for studying with background music, and participants could choose from these four responses: a) "To relax," b) "To concentrate on the learning material," c) "To get more energy," d) "Other." Again, participants who responded with "Other" were asked to provide some more details which were subsequently categorized by the candidate before the analysis. The second optional question was an open-ended query that asked participants to specify the type of music they commonly listen to while studying. No specific instructions were given, allowing participants to freely express their responses. The candidate then categorized these responses into music genres. Each music type was assigned a point in the corresponding category. If a participant mentioned multiple music types, each type received a point in the appropriate category.

Personality: A new addition to Study 2 was also the *Short 15-item Big Five Inventory* (BFI-S; Lang et al., 2011). BFI-S is a shorter version of the Big Five Inventory (BFI; John et al., 1991), a 44-item instrument measuring five dimensions of personality. The English version of the BFI-S has been validated and proven to be robust and a solid alternative to the much longer BFI, as it replicates the same five-factor structure (Lang et al., 2011), but is much more convenient to implement due to its shorter nature. Two researchers

translated the instrument from English to Slovene, addressed any disparities between the two renditions, and subsequently back-translated it into the original language. The instrument contains 15 items (four are reverse coded) or three per personality dimension. The factors measured by the BFI-S are: openness ($\omega = 0.770$; $\omega_{Slo} = 0.764$; $\omega_{Eng} = 0.661$), conscientiousness ($\omega = 0.700$; $\omega_{Slo} = 0.758$; $\omega_{Eng} = 0.503$), extraversion ($\omega = 0.846$; $\omega_{Slo} = 0.851$; $\omega_{Eng} = 0.852$), agreeableness ($\omega = 0.499$; $\omega_{Slo} = 0.476$; $\omega_{Eng} = 0.656$), and neuroticism ($\omega = 0.711$; $\omega_{Slo} = 0.703$; $\omega_{Eng} = 0.728$). All subscales had at least acceptable levels of internal consistency, except for agreeableness, which had poor reliability.

Participant characteristics: In addition to the standard demographic questions about the participants' gender, age, study program, academic year, and country of origin, two questions tailored to the study's focus were included. Considering the auditory independent variable, participants were asked whether they had any difficulties hearing. This query featured "yes" and "no" response options. If respondents selected "yes," they were invited to briefly describe what issues they might be experiencing. The second question inquired about potential formal musical training, once again offering "yes" and "no" response choices. If participants indicated "yes," they were further prompted to specify the duration of their musical training in terms of years.

3.5.2.5 Data collection

The collection of experimental data spanned from January to June 2023 and involved multiple on-site and online testing sessions. Students from different universities and on various levels were invited to partake in the study via a non-random selection process and were presented with two options – participating on-site in a group setting or participating individually online. Invitations were extended through email, social media posts, presentations, or by professors during or after lectures. Out of 307 participants, 252 were tested with the candidate present on-site, and 55 participated online.

In both versions, clear emphasis was placed on voluntary participation, and the participants were reminded at the beginning of the experiment (verbally or in written form) that they may withdraw at any point without explanation. Students read and confirmed an informed consent form before the start of the experiment. Participants were not offered any incentives for their involvement, but those who participated on-site had the opportunity to get some refreshments. Ethical approval for the research was obtained from the University of Primorska's Commission for Ethics in Human Subject Research before the first testing took place.

As was the case in Study 1, the whole survey/experiment was administered electronically through the 1ka.si online platform (Faculty of Social Sciences, University of Ljubljana, 2022). The whole procedure, which is depicted in Figure 10, lasted between 50 to 75 minutes, including providing instructions.

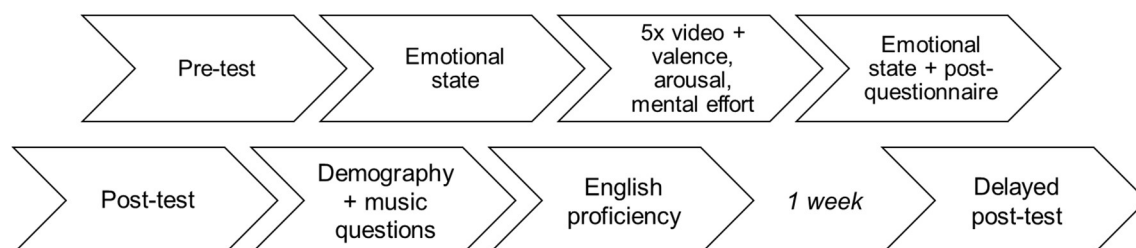


Figure 10: Experimental procedure of Study 2.

When the experiment was conducted on-site, each data collection session involved a group of 5 to 20 students. The testing sessions occurred in either a computer room equipped with faculty computers or a classroom where participants used their own laptops to watch videos and complete the survey. Most participants used headphones that were provided for them by the candidate, but some preferred to use their own headphones. While the candidate remained present throughout the experiment to address queries, participants completed the experiment individually and at their own pace.

The procedure for online participants was exactly the same as for the on-site participants, the only difference was in the form of the instructions provided. In the case of organized group testing sessions, the instructions were provided verbally by the candidate, while in the case of online testing, detailed instructions were written in the email that contained the link to the study.

All participants received uniform directions for every aspect of the study to ensure consistency of results across various sessions and between the different conditions. The survey included written instructions on how to respond to the survey and the parameters for playing videos (including volume level, playback without rewinding, fast-forwarding, or pausing, at standard speed, uniform video quality, and with subtitles deactivated). Prior to viewing the actual learning videos, a brief video test was administered to enable participants to adjust settings and identify any potential issues. The survey design also ensured that participants completed all items within a specific scale before progressing to the next section of the questionnaire, minimizing the possibility of missing data.

Since the candidate did not have control over the testing environment of online participants, additional directions were given to them regarding the context in which they were to perform the study. Online participants were instructed to allocate at least 90 minutes to the experiment and to finish it in one sitting without any breaks. They were also instructed to use a computer and headphones for participating and to make sure to be in a quiet room without distractions.

Experimental conditions were randomly assigned to participants in each session. For the on-site testing, the candidate prepared paper slips with links corresponding to each condition, based on the number of participants. These links were placed in a bag, and participants drew them randomly, ensuring an equal distribution across all conditions. For online participants, the website nimblelinks.com was used to randomly redistribute each click to one of the three experimental conditions.

At the conclusion of the survey, participants were prompted to provide their student email and a 6-digit identification code. This code was derived from the first two letters of their mother's name, the day of their birth month, and the first two letters of their birthplace. This information was solely used to send participants a link for the delayed post-test and to connect the data from both testing sessions. The contact information and identification code were promptly deleted when the data was connected with the proper participant.

Seven days after each participant completed the main experiment session, they were sent an email thanking them for their participation, prompting them to respond to the same knowledge test again, and informing them about how to reach the candidate and when the data will be available.

3.5.2.6 Data analysis

Analysis and outcomes of the data were conducted through Microsoft Excel and the open-source software jamovi (The jamovi project, 2022). Descriptive statistics were calculated for all variables, including the mean, standard deviation, the lowest and highest responses/results, kurtosis, and skewness. Before proceeding to further analyses, boxplots were examined for the presence of outliers, and assumptions of normality and homogeneity of variances were assessed through Shapiro-Wilk and Levene's (or Box's) tests, respectively. As was done in the case of the first experiment, the results of the homogeneity and normality tests are presented in Appendix 20. However, when only the normality assumption was not met, parametric tests were still

used, as the violation of the normality assumption had a minimal effect when a sample was large (Field, 2018).

The three groups were firstly compared on several characteristics before the introduction of the independent variable with one-way analysis of variance (ANOVA) for ordinal or continuous variables and with χ^2 for nominal variables such as gender, study level, study program, and country of origin. Control variables were also evaluated for differences among the groups depending on their educational field (wood science, engineering, and forestry-related or not), using either Student's or Welch's *t*-tests. Open-ended responses in the music-listening section were categorized by the candidate by identifying and assigning them to specific content themes. If a response contained multiple themes, it was categorized into all relevant categories, with each relevant category receiving one point.

The main analysis of differences between experimental groups in dependent variables mimicked the strategy used in Study 1. First, outcomes were compared using multiple Fisher's ANOVAs, which assume equal variances. When assumption tests revealed that the variances were not equal, Welch's ANOVA was conducted instead. In case a significant effect was observed, post-hoc tests were also performed (Tukey when variances were equal and Games-Howell post-hoc test when they were not). Next, groups were compared utilizing multiple analyses of covariance (ANCOVA) to control for the influence of potentially confounding variables. ANCOVAs were also used when analysing data that was collected multiple times during the experimental procedure and included a baseline measure, such as the PANAVA-KS, valence, and activation level variables. Lastly, multivariate analysis of covariance (MANCOVA) was also used in cases of multiple related dependent variables, to see whether the data has additional nuances not detectable with multiple ANCOVAs. The same variables as in the previous experiment were included, with the novelty being the Big 5 personality components and the variable that was shown to vary between groups before the video watching. Together, eleven (or ten in some cases) variables were added as covariates: personality traits (openness, conscientiousness, extroversion, agreeableness, and neuroticism), the three PANAVA-KS baseline measures (except cases where participants' valence and activation level were assessed using two single items, for which the same two baseline measures will be applied instead of the PANAVA-KS baseline measures), prior interest, tested prior knowledge and tested English proficiency. While the number of covariates is quite high, it still falls within the guideline advising there should be at least 10 events

(subjects) per predictive variable to help mitigate the risk of overfitting (Harrell et al., 1984).

To address the greater number of comparisons and minimize the heightened risk of Type I errors (Colman, 2014), a Bonferroni correction was applied. However, due to the exploratory nature of the research and the small expected effect of the independent variable, marginally significant differences with p -values of less than 0.10 will also be highlighted and interpreted with caution.

Power calculations with the G*Power software (Faul et al., 2007) based on our sample size, an alpha level of 0.05, and a desired power of $1 - \beta = 0.80$ indicated that the study has the ability to detect a small to medium effect size (Cohen's $f = 0.18$), meaning that our research design is adequately powered to detect meaningful distinctions between the three groups. On the other hand, the reduced sample size of 118 participants during the delayed phase allows us to detect a medium to large effect size (Cohen's $f = 0.29$).

At the end, comparisons based on English proficiency, study program, and personality were also conducted.

3.5.3 Results and interpretation

This chapter will begin with a description of the sample and a comparison between the three groups on variables that could affect the results. Then, the results will be divided into subchapters based on types of variables, namely emotional outcomes, cognitive outcomes, and learning. As was done in Study 1, subchapters will include results of ANOVAs, MANCOVAs, and singular ANCOVAs.

To address the issue of multiple comparisons and minimize the risk of Type I errors, a Bonferroni correction was implemented for all analyses (Colman, 2014). Specifically, for the initial group comparison before the intervention, we set the α level at 0.003 (0.05/18), while for the evaluation of group differences concerning dependent variables, an α level of 0.002 (0.05/32) was employed. Nevertheless, marginally significant results with the α level of up to 0.10 will be also pointed out due to the exploratory nature of the study.

3.5.3.1 Groups' description and comparison

Before delving into the main results of Study 2, initial analyses were performed to investigate potential variations between participants in different groups, as these

differences could influence the changes observed in our dependent variables. Initially, an overview of the entire study sample will be provided, followed by descriptive statistics and comparisons of control variables divided by experimental groups.

Additionally, participants will also be divided based on their study program – whether their field of study is related to wood science, engineering, and forestry or not. Specifically, the first group will consist of participants from educational fields falling under the Klasius-P-16 Engineering, manufacturing and construction and Agriculture, forestry, fisheries, and veterinary ($N = 106$), while the second group will consist of all other participants (students of all other educational fields, $N = 194$). This was done because we predicted there being significant differences between participants in the two groups in variables, such as prior knowledge, experience, and interest in the topic.

Looking at the whole sample (students from educational fields both related and not related to the topic of the learning videos and students whose English is their native and non-native language), participants evaluated their level of prior knowledge about wood as a building material somewhat low ($M = 3.15$, $Mdn = 3$, $SD = 1.49$), which aligned with the findings from the pre-test, where the average score was 2.12 ($Mdn = 2$, $SD = 1.73$) out of 8. On average, they evaluated that 65% of the learning video content was new knowledge for them ($M = 65.12$, $Mdn = 75$, $SD = 26.51$). Most participants reported having (very) rarely worked with wood in the past ($M = 3.34$, $Mdn = 3$, $SD = 1.66$), but were somewhat interested in learning about the topic ($M = 4.61$, $Mdn = 5$, $SD = 1.71$). On average, participants were at least somehow fluent in 3 languages ($M = 3.32$, $Mdn = 3$, $SD = 1.43$) and rated their English comprehension as somewhat high ($M = 5.50$, $Mdn = 6$, $SD = 1.43$), which was confirmed by their relatively high average score on the English vocabulary test ($M = 71.02$, $Mdn = 70.00$, $SD = 14.01$).

Table 75 displays descriptive statistics for these variables, categorized based on whether participants are pursuing educational fields related to wood science, forestry, and engineering or not, while Table 76 presents the comparison between the two groups, together with assumption checks.

Table 75: Learners' characteristics and descriptive statistics for variables before watching the videos divided by educational field

	<i>M</i>		<i>SD</i>		Min–Max		Skewness		Kurtosis	
	W	NW	W	NW	W	NW	W	NW	W	NW
SPK	3.95	2.70	1.44	1.34	1–7	1–7	0.06	0.75	–0.44	0.31
TPK	3.23	1.51	1.86	1.29	0–8	0–6	0.35	0.73	–0.51	0.24
NK	50.19	73.5 7	27.80	21.6 2	0– 100	0– 100	0.01	–1.04	–1.07	0.64
PE	4.40	2.75	1.75	1.30	1–7	1–7	– 0.10	0.90	–1.01	0.57
PI	5.68	4.03	1.46	1.54	1–7	1–7	– 1.39	–0.27	1.65	–0.76
Lan	2.86	3.57	1.35	1.41	1–9	1–10	1.19	1.06	3.41	2.71
SEP	5.45	5.52	1.59	1.35	1–7	2–7	– 1.14	–0.49	0.74	–0.76
TEP	72.44	70.2 4	16.00	12.7 7	40– 100	43.7 5– 100	0.01	0.28	–0.99	–0.45
Val ^b	5.75	5.82	1.37	1.55	1–9	1–9	– 0.47	–0.32	1.11	0.28
AL ^b	5.02	5.31	1.57	1.73	1–9	1–9	– 0.45	–0.15	0.39	0.08
PA ^b	4.04	4.16	1.11	1.18	1.25 – 6.50	1–7	– 0.19	–0.02	–0.01	–0.46
NA ^b	3.32	3.15	1.17	1.30	1–6	1–7	0.09	0.24	–0.44	–0.54
VA ^b	4.90	4.92	1.16	1.21	1–7	1–7	– 0.70	–0.33	0.79	–0.31
O	5.10	5.26	1.17	1.14	2–7	2.33 –7	– 0.47	–0.47	–0.22	–0.36
C	5.12	5.11	0.97	0.92	2.33 –7	2.33 –7	– 0.33	–0.13	0.29	–0.37
E	4.25	4.61	1.36	1.45	1– 6.67	1–7	– 0.29	–0.28	–0.14	–0.58
A	5.13	4.89	0.93	1.04	3.33 –7	2–7	0.02	–0.36	–0.67	–0.03

N	3.91	4.12	1.32	1.25	1–7	1–7	– 0.06	–0.18	–0.32	–0.52
MU	3.07	2.78	1.65	1.52	1–6	1–6	0.21	0.40	–1.19	–0.97
Age	23.38	22.2 4	5.36	4.75	19– 48	17– 56	2.21	3.51	5.39	17.60

Note. W – participants from fields of education related to wood science, engineering, and forestry (KLASIUS-P–16 groups: Engineering, manufacturing and construction and Agriculture, forestry, fisheries and veterinary), NW – participants from other fields of education; SPK – subjective prior knowledge, TPK – tested prior knowledge, NK – new knowledge, PE – prior experience, PI – prior interest, Lan – number of spoken languages, SEP – subjective English proficiency, TEP – tested English proficiency, Val – valence, AL – activation level, ^b – baseline, PA – positive activation, NA – negative activation, VA – valence, O – openness, C – conscientiousness, E – extroversion, A – agreeableness, N – neuroticism, MU – music listening

Given that certain variables did not adhere to the assumptions of normality (as determined by the Shapiro-Wilk test) and homogeneity of variances (indicated by Levene's test), we opted for Welch *t*-tests for all variables instead of Student's *t*-tests. Compared to the Student's *t*-test, the Welch's *t*-test provides better control over Type I error rates when the assumption of homogeneity of variance is not met, but when the variances are equal, the Welch *t*-test performs similarly to the Student's *t*-test (Delacre et al., 2017). Moreover, when compared to the widely used non-parametric alternative, the Mann-Whitney *U* test, the Welch *t*-test exhibits a similar level of Type I error control when variances are equal but surpasses the *U* test when variances are unequal (Ruxton, 2006; Zimmerman and Zumbo, 1993).

Consistent with our expectations, participants in educational fields closely related to the video content exhibited markedly greater prior knowledge, both self-assessed and assessed through testing, as well as more substantial experience working with wood and a higher level of interest in the subject (Table 76). In addition, individuals within this group reported a significantly lower percentage of the video content that was new information to them compared to students from other educational fields. This effect size for all these variables was large ($d > 0.90$). The same group also reported a lower number of spoken languages. There were also marginally significant differences in the level of extroversion and agreeableness between the two groups. No other differences were observed.

Table 76: Comparisons of learners' characteristics before watching the videos divided by educational field using Welch's *t*-tests

	Welch' <i>t</i> -tests				Assumption checks	
	<i>t</i>	<i>df</i>	Mean difference [95% CI]	<i>d</i> [95% CI]	<i>W</i>	<i>F</i> [†]
SPK	7.38***	204.04	1.25 [0.92–1.59]	0.90 [0.64–1.17]	0.96***	0.03
TPK	8.46***	161.71	1.72 [1.32–2.12]	1.07 [0.85–1.41]	0.96***	18.69***
NK	−7.51***	175.47	−23.38 [− 29.53—17.24]	−0.94 [−1.24—0.70]	13.87***	0.97***
PE	8.48***	169.51	1.64 [1.26–2.03]	1.07 [0.83–1.39]	0.97***	22.09***
PI	9.18***	225.30	1.65 [1.30–2.01]	1.10 [0.81–1.37]	0.95***	1.75
Lan	−4.31***	224.34	−0.71 [−1.04—0.39]	−0.52 [−0.76—0.27]	0.92***	1.26
SEP	−0.37	188.18	−0.07 [−0.43–0.29]	−0.05 [−0.28–0.19]	0.87***	1.31
TEP	1.22	179.31	2.20 [−1.36–5.76]	0.15 [−0.08–0.39]	0.98**	13.42***
Val ^b	−0.46	239.06	−0.08 [−0.42–0.26]	−0.05 [−0.29–0.18]	0.96***	3.20
AL ^b	−1.50	234.76	−0.30 [−0.68–0.09]	−0.18 [−0.41–0.06]	0.98***	1.16
PA ^b	−0.87	228.92	−0.12 [−0.39–0.15]	−0.10 [−0.34–0.13]	1.00	1.41
NA ^b	1.18	236.03	0.17 [−0.12–0.46]	0.14 [−0.10–0.38]	0.98**	2.87
VA ^b	−0.12	223.32	−0.02 [−0.30–0.26]	−0.01 [−0.25–0.22]	0.97***	0.47
O	−1.13	210.96	−0.16 [−0.44–0.12]	−0.14 [−0.38–0.10]	0.97***	0.09
C	0.11	205.50	0.01 [−0.21–0.24]	0.01 [−0.22–0.25]	0.98***	0.09
E	−2.12*	227.13	−0.36 [−0.69—0.03]	−0.25 [−0.49—0.01]	0.98***	1.18

A	1.99*	237.57	0.23 [0.00–0.46]	0.24 [–0.01–0.47]	0.99*	0.81
N	–1.33	205.87	–0.21 [–0.52–0.10]	–0.16 [–0.40–0.08]	0.99*	0.22
MU	1.46	201.98	0.28 [–0.10–0.66]	0.18 [–0.06–0.42]	0.92***	0.70
Age	1.82	194.82	1.14 [–0.09–2.36]	0.22 [–0.01–0.47]	0.69***	3.07

Note. *CI* – confidence interval, *d* – effect size; [†] *df*1 = 1, *df*2 = 298; * *p* < 0.05, ** *p* < 0.01, *** *p* < .001; SPK – subjective prior knowledge, TPK – tested prior knowledge, NK – new knowledge, PE – prior experience, PI – prior interest, Lan – number of spoken languages, SEP – subjective English proficiency, TEP – tested English proficiency, Val – valence, AL – activation level, ^b – baseline, PA – positive activation, NA – negative activation, VA – valence, O – openness, C – conscientiousness, E – extroversion, A – agreeableness, N – neuroticism, MU – music listening

Regarding demographical data, there were some significant differences in gender ($\chi^2(3, N = 300) = 26.106, p < .001$), as in the wood science, engineering, and forestry-related groups the ratio between genders was more equal (50 women, 53 men, 1 non-binary, 2 nondisclosed) than in the other group (148 women, 44 men, 1 non-binary and 1 nondisclosed) and in the country of origin ($\chi^2(21, N = 300) = 50.743, p < .001$), as the majority of participants in the first group came from Slovenia (79) and USA (15) and the majority of participants from the other group was from Slovenia (158), North Macedonia (15), USA (6), and Serbia (6). There were also some marginally significant differences in study level ($\chi^2(3, N = 300) = 6.991, p = 0.072$).

On the other hand, there were no significant differences between the two groups in music and hearing-related variables, such as having hearing difficulties ($\chi^2(1, N = 300) = 0.048, p = 0.826$), formal musical training ($\chi^2(1, N = 300) = 0.005, p = 0.945$), and motivation for listening to music while studying ($\chi^2(4, N = 300) = 5.652, p = 0.227$).

Next, the experimental groups will be described and compared. Participants were randomly allocated to the respective experimental groups, ensuring an even distribution of participants across groups based on their country, study year, and study program. However, there was less control over group allocation in the case of online participation, so the groups were not completely equal.

In the no background music condition, there was a total of 102 participants, with 71 being women, 30 men, and 2 who chose not to specify their gender. Of these, 80

participants were from Slovenia, five from the USA and from Serbia, and four from North Macedonia, with the remaining participants originating from various countries (Argentina, Bosnia and Herzegovina, Canada, Germany, Nigeria, Russia, Tunisia, and Ukraine).

In the calm background music condition, a total of 105 participants were involved. Among them, 62 were female, 35 were male, and 2 identified as non-binary, while the gender of the remaining 6 participants was unspecified. 81 participants who watched the videos with calm music were primarily from Slovenia, ten from North Macedonia, seven from the USA, and the remainder came from diverse countries (Belarus, Germany, Ghana, Norway, Poland, and Spain).

In the lively background music condition, there were 100 participants, consisting of 67 females, 32 males, and one participant who did not disclose their gender. Among these, 82 participants were from Slovenia, nine were from the USA, and the remaining participants represented various countries (Bosnia and Herzegovina, Canada, Croatia, Hungary, North Macedonia, Peru, Serbia, South Korea, and Thailand).

Table 77 presents the summary statistics for the control variables and characteristics of the learners divided by experimental group.

Table 77: Learners' characteristics and descriptive statistics for variables before watching the videos divided by group

	No music (<i>N</i> = 102)	Calm music (<i>N</i> = 105)	Lively music (<i>N</i> = 100)
Subjective prior knowledge			
<i>M</i> (<i>SD</i>)	2.98 (1.53)	3.11 (1.44)	3.35 (1.49)
Min–Max	1–7	1–7	1–7
Skewness	0.53	0.47	0.44
Kurtosis	–0.55	–0.30	–0.13
Tested prior knowledge			
<i>M</i> (<i>SD</i>)	2.03 (1.69)	2.18 (1.75)	2.14 (1.75)
Min–Max	0–7	0–7	0–8
Skewness	0.84	0.74	0.93
Kurtosis	0.31	0.06	0.71
New knowledge			
<i>M</i> (<i>SD</i>)	68.22 (26.71)	64.95 (25.01)	62.15 (27.71)
Min–Max	0–100	0–100	0–100
Skewness	–0.72	–0.69	–0.61

Kurtosis	−0.58	−0.37	−0.68
Prior experience			
<i>M (SD)</i>	3.05 (1.51)	3.50 (1.80)	3.46 (1.64)
Min–Max	1–7	1–7	1–7
Skewness	0.78	0.47	0.65
Kurtosis	0.04	−0.82	−0.53
Prior interest			
<i>M (SD)</i>	4.61 (1.69)	4.54 (1.74)	4.68 (1.71)
Min–Max	1–7	1–7	1–7
Skewness	−0.42	−0.30	−0.52
Kurtosis	−0.57	−0.96	−0.64
Number of spoken languages			
<i>M (SD)</i>	3.29 (1.34)	3.56 (1.68)	3.11 (1.20)
Min–Max	1–9	1–10	1–7
Skewness	1.05	1.04	0.43
Kurtosis	2.76	2.13	0.41
Subjective English proficiency			
<i>M (SD)</i>	5.41 (1.54)	5.53 (1.48)	5.56 (1.28)
Min–Max	1–7	1–7	2–7
Skewness	−0.90	−0.80	−0.60
Kurtosis	0.37	−0.12	−0.43
Tested English proficiency			
<i>M (SD)</i>	70.30 (13.71)	71.11 (14.74)	71.65 (13.66)
Min–Max	41.25–100	43.75–100	40.00–100
Skewness	0.28	0.16	0.16
Kurtosis	−0.66	−0.81	−0.43
Valence baseline			
<i>M (SD)</i>	5.58 (1.51)	5.97 (1.58)	5.80 (1.33)
Min–Max	1–9	2–9	1–8
Skewness	−0.00	−0.35	−0.88
Kurtosis	−0.13	−0.50	1.23
Activation baseline			
<i>M (SD)</i>	5.10 (1.79)	5.39 (1.68)	5.11 (1.53)
Min–Max	1–9	1–9	1–8
Skewness	−0.12	−0.23	−0.33

Kurtosis	−0.37	−0.12	−0.50
PA baseline			
<i>M</i> (<i>SD</i>)	4.04 (1.17)	4.11 (1.22)	4.15 (1.08)
Min–Max	1.50–6.50	1–7	1.25–6.50
Skewness	−0.03	−0.04	−0.11
Kurtosis	−0.35	−0.34	−0.28
NA baseline			
<i>M</i> (<i>SD</i>)	3.28 (1.34)	3.11 (1.32)	3.27 (1.08)
Min–Max	1–7	1–6.25	1.25–5.50
Skewness	0.12	0.30	0.04
Kurtosis	−0.53	−0.56	−0.67
VA baseline			
<i>M</i> (<i>SD</i>)	4.76 (1.22)	4.95 (1.24)	5.00 (1.08)
Min–Max	1–7	1.50–7	1.50–7
Skewness	−0.56	−0.35	−0.32
Kurtosis	0.26	−0.38	0.00
Openness			
<i>M</i> (<i>SD</i>)	5.18 (1.16)	5.20 (1.21)	5.22 (1.09)
Min–Max	2–7	2.33–7	2.33–7
Skewness	−0.45	−0.51	−0.42
Kurtosis	−0.19	−0.38	−0.39
Conscientiousness			
<i>M</i> (<i>SD</i>)	5.13 (0.97)	5.21 (0.90)	5.02 (0.92)
Min–Max	2.33–6.67	2.67–7	2.67–7
Skewness	−0.46	−0.05	−0.05
Kurtosis	0.26	−0.37	−0.22
Extroversion			
<i>M</i> (<i>SD</i>)	4.47 (1.38)	4.39 (1.42)	4.54 (1.49)
Min–Max	1–7	1–7	1–7
Skewness	−0.34	−0.25	−0.15
Kurtosis	−0.30	−0.63	−0.54
Agreeableness			
<i>M</i> (<i>SD</i>)	5.05 (0.98)	5.01 (1.03)	4.87 (0.97)
Min–Max	2.33–7	2–7	2.67–7
Skewness	−0.35	−0.56	0.07

Kurtosis	0.12	0.57	−0.69
Neuroticism			
<i>M (SD)</i>	4.06 (1.22)	3.85 (1.42)	4.23 (1.12)
Min–Max	1–6.67	1–7	1.67–7
Skewness	−0.31	0.04	0.02
Kurtosis	−0.10	−0.79	−0.41
Music listening			
<i>M (SD)</i>	2.88 (1.44)	3.02 (1.65)	2.75 (1.62)
Min–Max	1–6	1–6	1–6
Skewness	0.32	0.27	0.42
Kurtosis	−0.93	−1.13	−1.11
Age			
<i>M (SD)</i>	22.58 (5.35)	22.56 (4.79)	22.74 (4.82)
Min–Max	17–56	17–48	18–48
Skewness	3.38	2.76	2.59
Kurtosis	15.57	9.88	8.49

Note. PA – positive activation, NA – negative activation, VA – valence

To verify whether there were significant differences between participants in different groups even before the introduction of the independent variable, one-way analyses of variance were conducted with potentially confounding variables. Before that, assumptions for one-way ANOVAs were checked. Although a statistically significant Shapiro-Wilk test for normality is not a concern in large samples, as the distribution tends to approximate normality with sample sizes greater than 30, the violation of equal variances is a more significant issue, although not critical if sample sizes are equal (Field, 2018). Even though it is not necessary, we will report the alternative Welch's *F* statistic instead of the traditional Fisher's *F* statistic. Welch's *F* adjusts both the statistic and residual degrees of freedom, enhancing its robustness to address issues stemming from violations of the homogeneity of variance assumption (Field, 2018).

Table 78: One-way ANOVA comparisons of the learners' characteristics and variables before watching the videos between experimental groups

	ANOVA			Homogeneity test		Normality test	
	<i>F</i>	<i>p</i>	η^2p	<i>F</i>	<i>p</i>	<i>W</i>	<i>p</i>
Subjective prior knowledge	1.60	0.204	0.01	0.13	0.882	0.96	< .001
Tested prior knowledge	0.21	0.810	0.00	0.20	0.817	0.92	< .001
New knowledge	1.33	0.266	0.01	0.87	0.418	0.93	< .001
†Prior experience	2.49	0.086	0.01	3.82	0.023	0.94	< .001
Prior interest	0.16	0.849	0.00	0.44	0.644	0.94	< .001
†Number of spoken languages*	2.36	0.097	0.02	4.71	0.010	0.95	< .001
Subjective English proficiency*	0.31	0.733	0.00	1.89	0.153	0.88	< .001
Tested English proficiency**	0.23	0.791	0.00	0.58	0.560	0.98	0.001
Valence baseline	1.83	0.163	0.01	2.31	0.101	0.98	< .001
Activation baseline	1.02	0.362	0.01	1.03	0.357	0.98	< .001
PA baseline	0.22	0.804	0.00	0.70	0.495	1.00	0.486
NA baseline	0.60	0.550	0.00	2.39	0.093	0.98	0.002
VA baseline	1.12	0.328	0.01	1.17	0.313	0.98	< .001
Openness	0.04	0.965	0.00	0.45	0.637	0.97	< .001
Conscientiousness	1.04	0.355	0.01	0.48	0.618	0.99	0.009
Extroversion	0.31	0.737	0.00	0.51	0.603	0.98	0.001
Agreeableness	0.93	0.394	0.01	0.17	0.840	0.99	0.020
†Neuroticism	2.34	0.099	0.02	4.51	0.012	0.99	0.159
Music listening*	0.75	0.474	0.01	1.790	0.169	0.923	< .001
Age***	0.04	0.963	0.00	0.07	0.935	0.69	< .001

Note. $df_1 = 2$, $df_2 = 304$; * $df_1 = 2$, $df_2 = 297$; ** $df_1 = 2$, $df_2 = 296$; *** $df_1 = 2$, $df_2 = 299$; † Welch's instead of Fisher's *F* statistic is reported ($df_{2 \text{ prior experience}} = 201.97$; $df_{2 \text{ languages}} = 194.97$; $df_{2 \text{ neuroticism}} = 201.82$)

Table 78 displays the comparisons of control variables and characteristics of the learners between experimental groups. It can be noted that there are no statistically significant differences between the groups at the α level at 0.003. Nonetheless, it is important to highlight that three variables exhibited results that were marginally significant and had unequal variances, which led us to report their Welch's F statistic. Specifically, these variables were prior experience, the number of spoken languages, and neuroticism. For these particular variables, we conducted non-parametric Games-Howell post-hoc tests, while no post-hoc tests were done for other variables.

Table 79: Games-Howell post-hoc tests of three learners' characteristics and variables before watching the videos between experimental groups

			Mean difference	t	df	p
Prior experience	No music	Calm music	-0.45	-1.93	200.88	0.132
	No music	Lively music	-0.41	-1.85	197.93	0.156
	Calm music	Lively music	0.04	0.15	202.65	0.988
Number of spoken languages	No music	Calm music	-0.27	-1.27	188.74	0.411
	No music	Lively music	0.18	0.98	196.57	0.591
	Calm music	Lively music	0.45	2.17	179.58	0.079
Neuroticism	No music	Calm music	0.21	1.15	202.05	0.486
	No music	Lively music	-0.17	-1.06	199.11	0.541
	Calm music	Lively music	-0.39	-2.16	196.22	0.080

As can be seen in Table 79, post-hoc tests do not reveal any significant differences in control variables. The only marginally significant differences are between the participants in the calm and lively music conditions in spoken languages and neuroticism. However, these differences should not significantly affect the results. Additionally, there were no significant differences between the three groups in gender ($\chi^2(8, N = 302) = 6.979, p = 0.539$), study level ($\chi^2(6, N = 300) = 10.577, p = 0.102$) or program ($\chi^2(18, N = 299) = 12.223, p = 0.835$), and country of origin ($\chi^2(42, N = 306) = 49.439, p = 0.200$).

Participants were also compared in variables connected to hearing and music. No significant differences between the experimental groups were detected in having hearing difficulties ($\chi^2(2, N = 300) = 0.425, p = 0.808$), formal musical training ($\chi^2(2, N = 300) = 0.214, p = 0.899$), frequency of listening to music while studying (Table 78), and motivation for listening to music while studying ($\chi^2(8, N = 300) = 13.317, p = 0.101$).

In summary, it can be deduced that the groups were similar in fundamental characteristics and potential confounding factors prior to viewing the videos.

Finally, we will provide some information on the music listening habits of the whole sample while studying. As illustrated in Figure 11, more than a quarter of respondents (27.00%) indicated that they never listen to music during their study sessions. Additionally, around one-fifth of the participants reported that they either rarely (17.67%) or sometimes (20.67%) engage in studying with music. In contrast, nearly 30% of the participants have background music playing frequently during their study sessions, with 14.6% indicating they do so often and an equivalent percentage stating they do it very often. Lastly, a smaller group (5.33%) reported that they always have music playing while studying. These findings suggest a diverse range of music-listening behaviors among the study participants, with a substantial portion opting for background music during their study sessions.

The 219 participants who responded that they listen to music while studying at least rarely answered two additional questions. The first question aimed to understand their motivation behind listening to music during study sessions, and the second inquired about the type of music they preferred for studying.

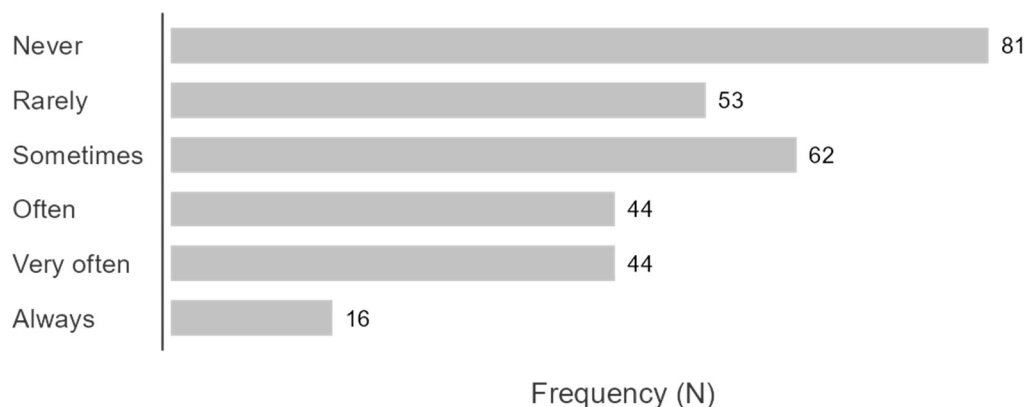


Figure 11: Frequency of studying with music in the background.

Regarding their motivation, the participants provided diverse responses. The three offered answers were the most popular and none of them stood out – 75 or 32.75% of the answers listed being able to relax as the main motivation, 71 or 31.00% to enhance their concentration on the learning material, and 66 or 28.82% to feel more energized and motivated. Six participants elaborated that they listen to music for all of the above reasons, depending on varying factors, such as their mood, the specific song playing, and the subject they were studying. For example, three participants (1.31%) specifically mentioned they listen to music only when studying math or doing calculations (“... *If I am*

studying a topic that requires some mathematics or calculations, I listen to is afrobeat fast rhythm or tempo. Aside this I rather don't have any music"). Some participants also mentioned using music to mask environmental noise (3.06% or 7 participants), prevent boredom (1.31% or 3 participants; "Because I'm doing the same thing for a long time and I'm already bored"), expose themselves to multiple stimuli (0.87% or 2 participants; "... so it's not complete silence"), and make the learning experience more enjoyable (0.87% or 2 participants; "Learning is more pleasant"), illustrating the multifaceted role of music in their study routines.

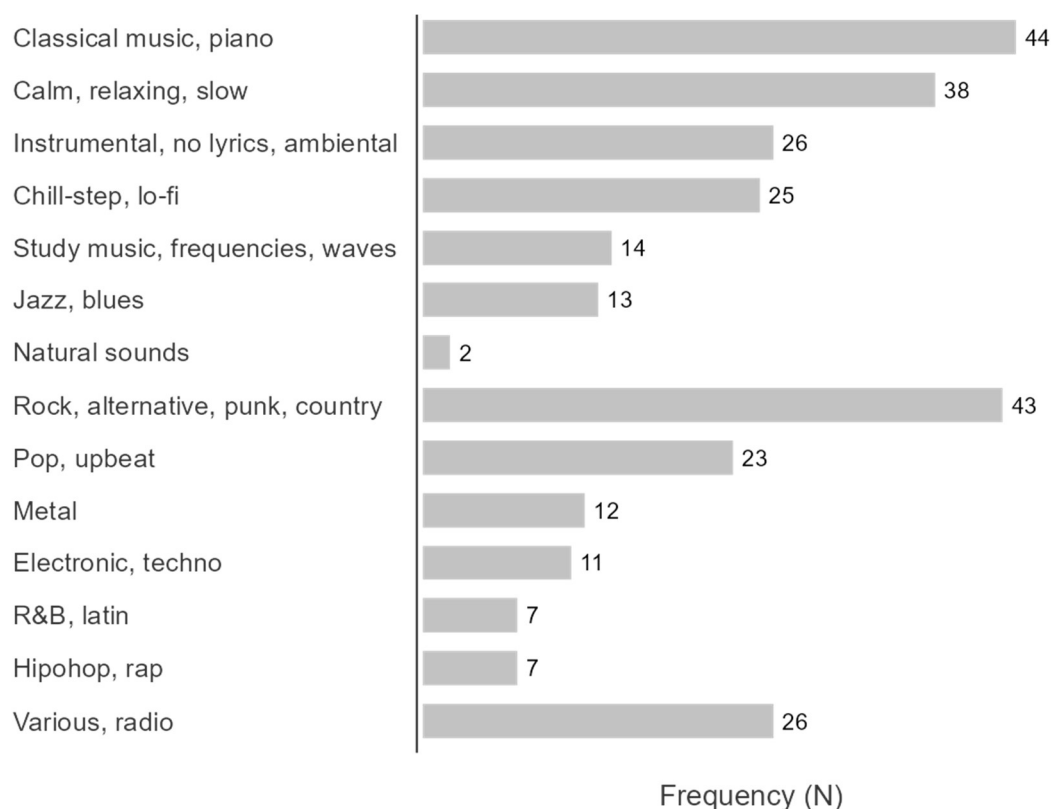


Figure 12: Frequency of music genres listened to during studying.

The same participants also reported what kind of music they listened to during their study sessions. The answers' distribution is illustrated in Figure 12, where the music genres or descriptors used (categorized by the candidate) are divided into calmer and livelier. 26 or 8.93% of the participants did not specify a particular music type and indicated that their choice varied depending on mood or involved listening to the radio, which plays a variety of different songs. Among the specified preferences, the majority (162 or 55.67% of the answers) leaned towards calmer and more relaxing music genres, while 103 or 35.40% of answers indicated a preference for livelier and more energetic music genres. Since the number of listed music genres varied between participants (e.g.,

some may have listed only one genre and others multiple different genres), conclusions on musical preferences cannot be made. Nonetheless, these results show that participants who study with music in the background have varied preferences regarding the type of music they listen to and it can depend also on other factors, such as their current mood.

3.5.3.2 Background music and video perception

First, we will present the perception of background music's influence on the participants. Due to the categorical nature of the variables, no statistical tests were made to compare the groups, so only a description of the frequency distribution will be provided.

Figure 13 presents the frequency of participants' answers from the most to least frequent descriptors. Among the provided answers, 26 participants chose the answer "*Other*" and provided further details, sometimes with more than one theme (i.e., "*Somehow it energized me, but also somehow distracted at the same time,*"), which were categorized in all relevant categories, making a total of 342 answers.

Participants predominantly perceived the additional background music as relaxing ($n = 100$), regardless of whether it was calm (64 or 20.85%) or lively (29 or 9.45%).

Following this, a significant number acknowledged not noticing any background music, as anticipated, particularly in the group without added music. However, six participants (1.75%) across the experimental groups with added background music also failed to register its presence.

14.66% (45) of participants found the added music detrimental to their learning experience, using words like distracting, disturbing, or annoying. They specifically highlighted its repetitiveness ("*It was very repetitive and annoying*") or noted its mismatch with the video content ("*...slightly disturbing because of the tone that doesn't match the character of the video - too relaxing, dreamy and almost sad*"). This sentiment was more pronounced for lively background music (29 or 9.45%) compared to calm background music (15 or 4.89%).

13.36% (41) of participants believed that the music had no impact on them, with nine (2.93%) from the calm and 14 (4.56%) from the lively music experimental group.

Conversely, 11 (3.58%) participants from both the calm and lively music groups felt the music helped them concentrate and pay attention to the video content. Additionally, 14 (4.56%) participants stated that lively music helped them feel more energized.

On the other hand, seven (2.28%) individuals watching videos with calm music reported that the music made them feel sleepy. This effect was also noted by two (0.65%) participants from the lively music condition. Additionally, two participants (0.65%) from both the calm and lively music conditions mentioned that the music made them feel more nervous, irritated, or anxious.

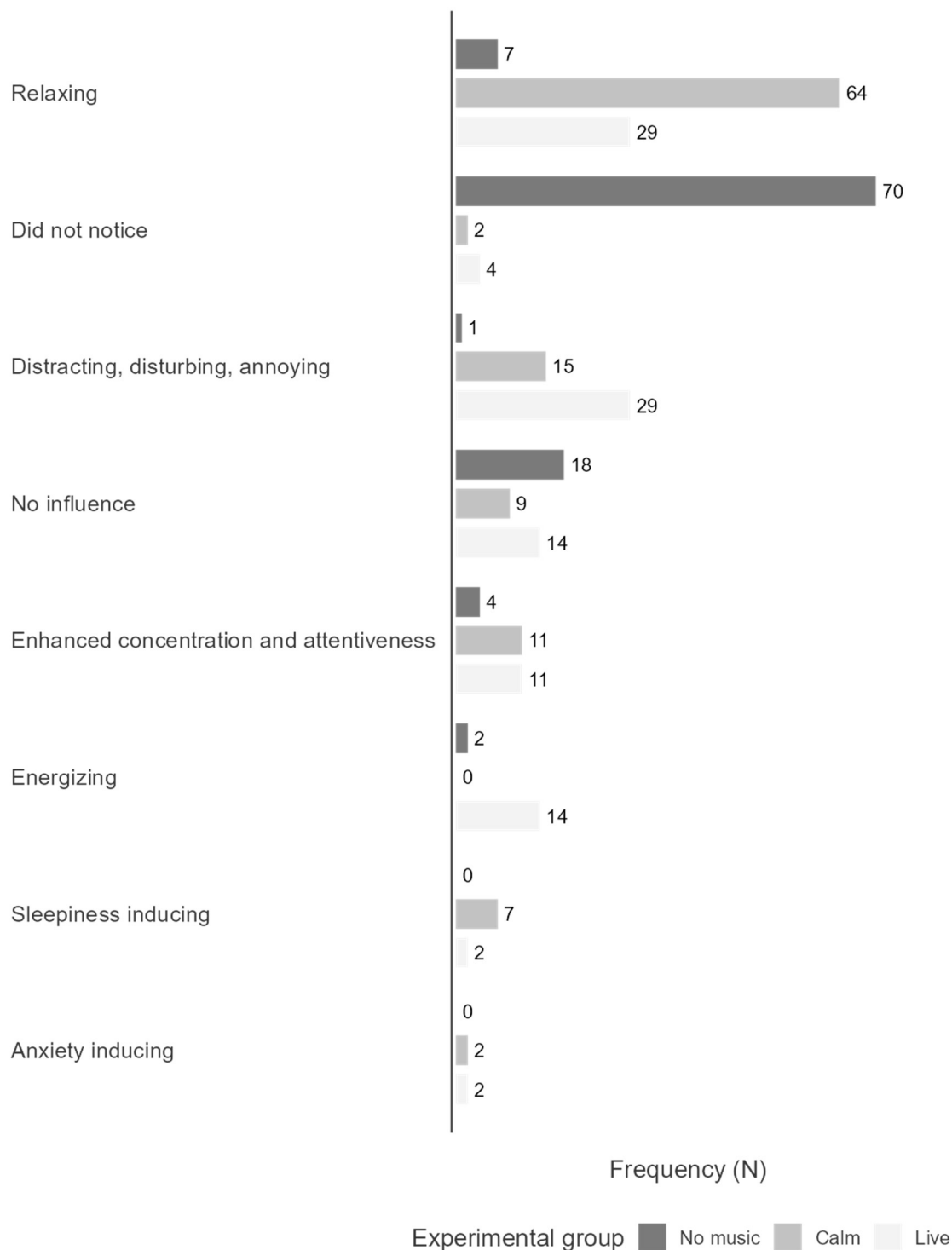


Figure 13: Background music influence perception answers frequency divided by group.

These responses were further organized into three overarching categories: the "Did not notice" and "No influence" responses were grouped under the "No effect" category; the responses indicating "Relaxing," "Enhancing concentration and attentiveness," and "Energizing" were categorized as "Positive effects"; and responses mentioning "Distracting," "Sleepiness," and "Anxiety-inducing" effects were placed in the "Negative effects" category. In summary, approximately a third (34.21% or 117) of the responses indicated no perceived effect, 42.52% (142) reported positive effects, and 16.96% (58) noted various negative effects. Figure 14 illustrates these categories, further segmented by experimental group.

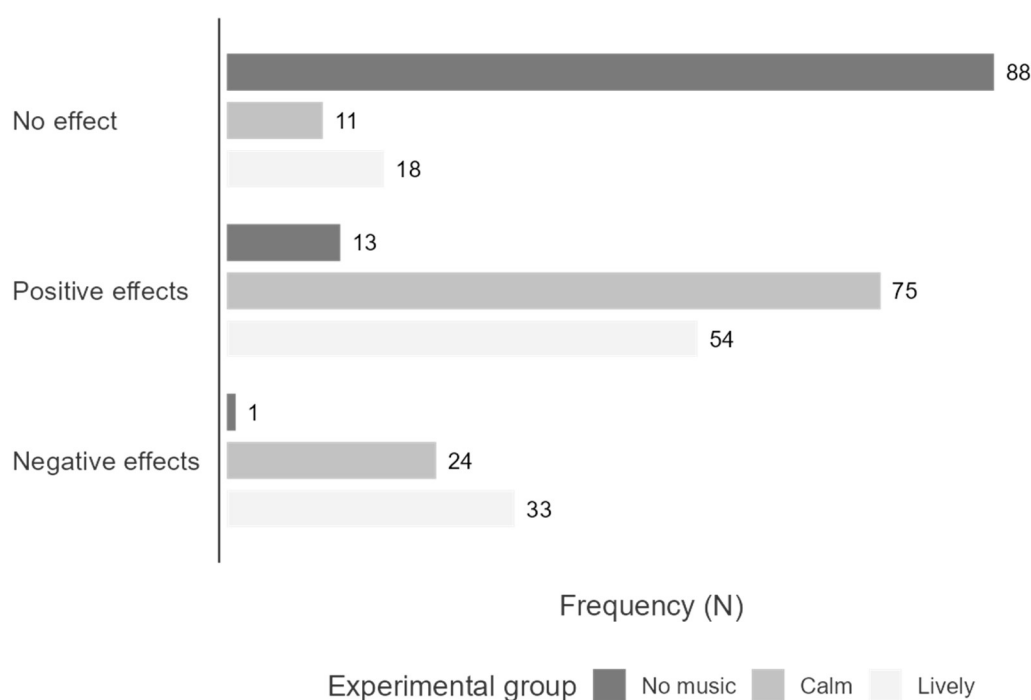


Figure 14: Background music influence perception categories frequency divided by group.

The answers of participants from the control group with no added background music predictably mostly (88 or 86.27%) fall under the "No effect" category, but there are 14 (13.73%) answers mentioning positive and negative effects. These responses were most likely related to the participants' beliefs of how background music affects them in general and not connected to the experiment, so they were disregarded.

In the calm music experimental group, a substantial portion (68.18% or 75) reported positive effects, while over a fifth (21.82% or 24) viewed calm music as adverse to their experience. About 10.00% (11) did not perceive either the music or its effects.

Similarly, in the lively music experimental group, almost half (49.52% or 54) felt the lively music positively influenced their experience, whereas 31.48% (33) reported negative effects. Approximately 17.14% (18) did not report any effects.

Analysing the categories, positive effects were more prevalent among participants in the calm music experimental group (75) compared to the lively music group (52). Conversely, individuals from the lively music group more frequently reported negative effects (33) or no effects (18) compared to the calm group (24 for negative and 11 for no effects).

These outcomes are intricately tied to the music selection for the experiment, implying that different chosen tracks might have yielded varied results. Another thing to note is that the influence of background music may have changed through time and repetitions, as highlighted by several participants: *“At the beginning, I was noticing the music but later it was as though there was no music. The music was distracting me at the beginning but later after the 3rd video, it was fine for me. I got used to it. At the end of one of the videos, I was asking myself if it had music in the background,”* or *“It was fine at first, but the repetition became annoying throughout the entire set of videos.”*

Next, participants were asked to rate the pleasantness and activation level of the videos as a whole (not music only). Both variables did not meet the assumption of normality but met the assumption of homogeneity of variances, so the parametric ANOVA was used. Table 80 displays the descriptive statistics of the two variables.

Table 80: Descriptive statistics for the two video perception variables divided by group

	No music	Calm music	Lively music
Perceived video activation level			
<i>M (SD)</i>	4.92 (1.41)	5.31 (1.38)	4.81 (1.39)
Min–Max	2–7	1–7	1–7
Skewness	–0.25	–1.12	–0.55
Kurtosis	–0.79	0.97	–0.24
Tested prior knowledge			
<i>M (SD)</i>	3.90 (1.61)	4.13 (1.40)	4.26 (1.38)
Min–Max	1–7	1–7	1–7
Skewness	–0.10	–0.16	–0.36
Kurtosis	–1.07	–0.57	–0.26

ANOVA comparisons revealed that there were no significant differences between the three groups in perceived pleasantness and activation level ($F(2,304) = 1.56$, $p = 0.213$, $\eta^2p = 0.01$) of the videos at the Bonferroni corrected p -value, but there was a marginally significant difference in the pleasantness of the videos ($F(2,304) = 3.73$, $p = 0.025$, $\eta^2p = 0.02$). Post-hoc comparisons for the latter variable are presented in Table 81. The results show that participants from the calm music experimental condition perceived the videos as significantly more pleasant than participants from the lively music condition.

Table 81: Post-hoc comparisons for Video pleasantness

Experimental groups		t	p	$p_{Bonferroni}$	Mean difference	d	95% CI
No music	Calm	-2.03	0.044	0.131	-0.39	-0.28	-0.56—0.01
No music	Lively	0.57	0.570	1.000	0.11	0.08	-0.20—0.36
Calm	Lively	2.59	0.010	0.030	0.20	0.09	0.09—0.64

Note. $df = 304$

To account for the influence of potentially confounding variables, a MANCOVA along with supplementary ANCOVAs were conducted (along with assumptions tests; Box's test and Shapiro-Wilk test for MANCOVA and Levene's and Shapiro-Wilk's tests for ANCOVAs). Eleven covariates were included in the model: pre-existing interest in the topic, prior knowledge, proficiency in English, initial emotional state (measured with the PANAVA-KS), and personality traits.

A MANCOVA produced a marginally significant effect of background music on perceived video pleasantness and energy level (Wilks' Lambda = 0.95, $F(4, 568) = 3.62$, $p = 0.006$; $\chi^2(6) = 2.69$, $p = 0.846$, $W = 0.98$, $p < .001$). Considering the similarity in individual MANCOVA outcomes and individual ANCOVAs, only the latter will be presented.

There was a marginally significant main effect for the video pleasantness ($F(2,285) = 3.76$, $p = 0.025$, $\eta^2p = 0.03$; $F(2, 296) = 0.86$, $p = 0.423$, $W = 0.98$, $p < .001$), but not for the perceived energy level of the videos ($F(2,285) = 1.64$, $p = 0.196$, $\eta^2p = 0.01$; $F(2, 296) = 2.06$, $p = 0.129$, $W = 0.99$, $p = 0.053$). Post-hoc comparisons for the first variable are presented in Table 82.

Table 82: ANCOVA post-hoc comparisons for Video pleasantness

Experimental groups		<i>t</i>	<i>p</i>	<i>p</i> _{Bonferroni}	Mean difference	<i>d</i>	95% <i>CI</i>
No music	Calm	-2.23	0.027	0.080	-0.42	-0.32	-0.60—0.04
No music	Lively	0.30	0.767	1.000	0.06	0.04	-0.24—0.33
Calm	Lively	2.49	0.013	0.040	0.48	0.36	0.07—0.65

Note. *df* = 285

Even after including covariates, the outcomes have not changed much, as individuals in the experimental condition exposed to calm music perceived the videos as notably more pleasant compared to those in the lively music condition. However, the calm music group also had marginally significantly higher ratings than the control group.

3.5.3.3 Emotional outcomes

This segment consists of several results, such as (differences in) affective states evaluated with three scales, interest in the subject of the videos, intrinsic motivation for video viewing, and learners' overall experience. Different measures of affective state exhibited a low to high positive correlation ($0.229 < r < 0.522$, $p < .001$) and a moderate negative correlation with the negative activation scale ($-0.338 < r < -0.485$, $p < .001$; Appendix 19).

Differences in affective state

As in the first experiment, participants' emotional states were evaluated through three scales: the Positive Activation, Negative Activation, and Valence Short Scale (PANAVA-KS) and two single-item scales gauged participants' activation level and valence. Participants completed the PANAVA-KS both before and after they viewed all the videos, while the single-item scales were administered before the first video to establish a baseline and after each of the subsequent five videos, making it a total of six administrations. Group comparisons were made using ANCOVAs to include the baseline measures. First, results related to the PANAVA-KS will be presented (Table 83), followed by findings on the two single-item scales.

Table 83: Descriptive statistics of PANAVA-KS values and change score divided by group

	No music	Calm music	Lively music
Positive activation baseline			
<i>M (SD)</i>	4.04 (1.17)	4.11 (1.22)	4.15 (1.08)
Min–Max	1.50–6.50	1–7	1.25–6.50
Skewness	–0.03	–0.04	–0.11
Kurtosis	–0.35	–0.34	–0.28
Positive activation after videos			
<i>M (SD)</i>	3.76 (1.23)	3.87 (1.34)	3.73 (1.13)
Min–Max	1–6.50	1–7	1–6.50
Skewness	–0.05	0.17	0.01
Kurtosis	–0.25	0.00	–0.17
Positive activation change score			
<i>M (SD)</i>	–0.28 (1.16)	–0.17 (1.19)	–0.17 (0.98)
Min–Max	–4–2.25	–3.25–4.25	–3–2
Skewness	–0.42	0.82	–0.02
Kurtosis	0.52	2.07	0.04
Negative activation baseline			
<i>M (SD)</i>	3.28 (1.34)	3.11 (1.32)	3.27 (1.08)
Min–Max	1–7	1–6.25	1.25–5.50
Skewness	0.12	0.30	0.04
Kurtosis	–0.53	–0.56	–0.67
Negative activation after videos			
<i>M (SD)</i>	3.11 (1.22)	2.75 (1.17)	3.18 (1.15)
Min–Max	1–5.50	1–6.25	1–5.75
Skewness	–0.02	0.60	0.07
Kurtosis	–0.79	0.32	–0.47
Negative activation change score			
<i>M (SD)</i>	–0.17 (1.08)	–0.35 (0.99)	–0.09 (0.99)
Min–Max	–6–2.25	–4.25–1.50	–3–2.50
Skewness	–1.45	–1.20	–0.26
Kurtosis	7.89	2.73	1.16
Valence baseline			
<i>M (SD)</i>	4.76 (1.22)	4.95 (1.24)	5.00 (1.08)

Min–Max	1–7	1.50–7	1.50–7
Skewness	–0.56	–0.35	–0.32
Kurtosis	0.26	–0.38	0.00
Valence after videos			
<i>M (SD)</i>	4.59 (1.26)	4.90 (1.28)	4.76 (1.12)
Min–Max	1–7	1.50–7	1.50–7
Skewness	–0.47	–0.62	–0.54
Kurtosis	0.41	0.35	0.04
Valence change score			
<i>M (SD)</i>	–0.17 (1.28)	–0.04 (1.16)	–0.23 (0.95)
Min–Max	–4–5.50	–5–4.50	–3–3
Skewness	0.94	–0.01	–0.17
Kurtosis	4.81	4.98	0.88

Note. change score – baseline measure subtracted from the second measure

Before making comparisons between groups, we conducted separate paired samples *t*-tests to assess the changes in the PANAVA-KS subscales within each of the three groups to see changes in affective states after the intervention (Table 84).

Table 84: Baseline and post-intervention differences in the PANAVA-KS measures for the three experimental groups separately

	<i>t</i>	<i>p</i>	<i>Mean difference</i>	<i>95% CI</i>	<i>d</i>	<i>95% CI</i>
No music group						
PA	2.435	0.017	0.279	0.05–0.51	0.241	0.04–0.44
NA	1.588	0.086	0.169	–0.04–0.38	0.157	–0.04–0.35
VA	1.353	0.036	0.127	–0.08–0.42	0.134	–0.06–0.33
Calm music group						
PA	2.105	0.038	0.245	0.01–0.48	0.205	0.01–0.40
NA	3.638	< .001	0.352	0.16–0.54	0.355	0.16–0.55
VA	0.378	0.706	0.043	–0.18–0.27	0.037	–0.15–0.26
Lively music group						
PA	4.276	< .001	0.417	0.22–0.61	0.428	0.22–0.63
NA	0.912	0.364	0.090	–0.11–0.29	0.091	–0.11–0.29

VA	2.462	0.016	0.235	0.05–0.42	0.246	0.05–0.45
----	-------	-------	-------	-----------	-------	-----------

Note. *CI* – confidence interval, *d* – effect size, *df*_{No music} = 101, *df*_{Calm music} = 104, *df*_{Lively music} = 99; PA – positive activation, NA – negative activation, VA – valence

In the control group, there was no statistically significant difference at the Bonferroni level, but all three were marginally significant, with all three post-intervention measures being lower than their baseline measure ($M_{PA-baseline} = 4.04$, $SD_{PA-baseline} = 1.17$; $M_{PA-after} = 3.77$, $SD_{PA-after} = 1.23$; $M_{NA-baseline} = 3.28$, $SD_{NA-baseline} = 1.34$; $M_{NA-after} = 3.11$, $SD_{NA-after} = 1.22$; $M_{VA-baseline} = 4.76$, $SD_{VA-baseline} = 1.22$; $M_{VA-after} = 4.59$, $SD_{VA-after} = 1.26$).

Conversely, in the group who watched videos with calm music, there was a statistically significant decrease in negative activation ($M_{NA-baseline} = 3.11$, $SD_{NA-baseline} = 1.32$; $M_{NA-after} = 2.76$, $SD_{NA-after} = 1.17$) and a marginally significant decrease in positive activation ($M_{PA-baseline} = 4.11$, $SD_{PA-baseline} = 1.22$; $M_{PA-after} = 3.87$, $SD_{PA-after} = 1.34$), but no difference in valence ($M_{VA-baseline} = 4.95$, $SD_{VA-baseline} = 1.24$; $M_{VA-after} = 4.91$, $SD_{VA-after} = 1.28$).

Finally, in the lively music group, the post-intervention positive activation score was moderately and significantly lower than the baseline measure ($M_{PA-baseline} = 4.15$, $SD_{PA-baseline} = 1.08$; $M_{PA-after} = 3.73$, $SD_{PA-after} = 1.13$), and the difference in valence also approached significance ($M_{VA-baseline} = 5.00$, $SD_{VA-baseline} = 1.08$; $M_{VA-after} = 4.76$, $SD_{VA-after} = 1.12$). The difference in negative effect in this group was negligible ($M_{NA-baseline} = 3.27$, $SD_{NA-baseline} = 1.08$; $M_{NA-after} = 3.18$, $SD_{NA-after} = 1.15$).

In summary, the results suggest that the interventions had a slightly different effect on participants' affective states.

Prior to comparing differences between the three groups, assumption checks for ANCOVA were conducted, including Levene's and Shapiro-Wilk's tests. While there were some violations of the normality assumption ($W_{PA} = 0.996$, $p_{PA} = 0.660$; $W_{NA} = 0.981$, $p_{NA} < .001$; $W_{VA} = 0.963$, $p_{VA} < .001$), the homogeneity of variances assumption was met in all cases ($F_{PA}(2, 304) = 2.116$, $p_{PA} = 0.122$; $F_{NA}(2, 304) = 0.543$, $p_{NA} = 0.582$; $F_{VA}(2, 304) = 0.716$, $p_{VA} = 0.489$), so three ANCOVAs were performed, with the second measure of each PANAVA-KS subscale as the dependent variable and the baseline measure as a covariate.

Between the three groups, there were no significant differences in positive activation ($F(2, 303) = 0.624$, $p = 0.536$, $\eta^2p = 0.004$) and valence ($F(2, 303) = 1.240$, $p = 0.291$,

$\eta^2p = 0.008$). However, there was a marginally significant effect in negative activation ($F(2, 303) = 3.681, p = 0.026, \eta^2p = 0.024$), so a post-hoc comparison was conducted. Table 85 shows a small to moderate difference in negative activation between the group with the calm and lively music.

Table 85: Post-hoc comparisons for Negative activation

Experimental groups		<i>t</i>	<i>p</i>	<i>p</i> _{Bonferroni}	Mean difference	<i>d</i>	95% <i>CI</i>
No music	Calm	1.99	0.047	0.142	-0.39	0.28	0.01–0.55
No music	Lively	-0.60	0.551	1.000	0.11	-0.08	-0.36–0.19
Calm	Lively	-2.58	0.010	0.031	0.20	-0.36	-0.64–0.09

Note. *df* = 303

Although Hypothesis 9 predicted differences in positive activation between the groups, our results indicate that background music primarily affects participants' negative activation. Specifically, participants in the calm music group experienced a greater reduction in negative activating emotions compared to those in the lively music group. While we expected lively music to positively enhance learners' arousal by amplifying emotions like excitement, happiness, and enthusiasm, the actual outcome was that calm music significantly lowered participants' negative arousing emotions such as worry, nervousness, and anger, thus calming them. Although our hypothesis was not supported, our results demonstrate that background music in educational videos impacts learners' emotions. Additionally, it highlights that not just the presence of music, but the type of music matters, as the effect on learners' emotions was significantly more pronounced in the calm music group compared to the lively music group.

Our findings are consistent with research indicating that high-arousal positive valence songs provided by researchers have a much lower impact on participants' feelings of joy and physiological activity compared to songs chosen by the participants themselves (Lynar et al., 2017), emphasizing the importance of musical preference. While researcher-selected lively and energetic music may not have a strong effect on students, possibly explaining the lack of differences in positive activation in our study, the low-arousal music chosen by researchers was still most effective in relaxing participants. Furthermore, our result that calm music videos led to a more significant decrease in negative activation supports the finding that music has the greatest impact on participants experiencing higher stress levels (Lynar et al., 2017).

As was done in Study 1, we conducted a MANCOVA and three additional ANCOVAs to examine the potential impact of other confounding variables (in addition to the baseline PANAVA-KS measures, prior interest, assessed prior knowledge, evaluated English proficiency, and the five personality traits were added).

While neither of the MANCOVA assumptions was satisfied in this case ($\chi^2(110) = 200.95$, $p < .001$, $W = 0.76$, $p < .001$), as mentioned earlier, the significant Shapiro-Wilk normality test and Box's test do not present an issue in large and comparable samples (Field, 2018). Hence, we can proceed with MANCOVA. The MANCOVA results indicated that the inclusion of background music has a marginally significant impact on the affective state of participants (Wilks' Lambda = 0.87, $F(20, 554) = 2.08$, $p = 0.004$).

Upon proceeding with individual tests, assumption checks confirmed the appropriateness of ANCOVA in all three cases (positive activation scale: $W = 1.00$, $p = 0.562$; $F(2, 296) = 2.11$, $p = 0.123$; negative activation scale: $W = 0.99$, $p = 0.013$; $F(2, 296) = 0.96$, $p = 0.384$; valence scale: $W = 0.96$, $p < .001$; $F(2, 296) = 0.33$, $p = 0.722$).

Table 86: ANCOVA post-hoc comparisons for Negative activation

Experimental groups		<i>t</i>	<i>p</i>	<i>p</i> _{Bonferroni}	Mean difference	<i>d</i>	95% <i>CI</i>
No music	Calm	2.08	0.039	0.116	0.26	0.28	0.01–0.58
No music	Lively	−0.16	0.872	1.000	−0.02	−0.08	−0.31–0.26
Calm	Lively	−2.21	0.028	0.084	−0.29	−0.36	−0.60–0.03

Note. *df* = 285

Despite controlling for covariates, the results remained unchanged, with no significant differences between the two groups in positive activation ($F(2, 285) = 0.54$, $p = 0.585$, $\eta^2p = 0.00$) and valence ($F(2, 285) = 1.59$, $p = 0.205$, $\eta^2p = 0.01$), but a marginally significant effect in negative activation ($F(2, 285) = 3.07$, $p = 0.048$, $\eta^2p = 0.02$). As can be seen from Table 86, there might be a small and marginally significant difference between the calm and lively music groups in negative activation.

Next, participants' affective state was assessed through two one-item questions regarding their valence and activation before and after each video. Descriptive statistics for all measures are depicted in Table 87.

Table 87: Descriptive statistics of activation level and valence measurements and change score divided by group

	No music	Calm music	Lively music
Activation level baseline			
<i>M (SD)</i>	5.10 (1.79)	5.39 (1.68)	5.11 (1.53)
Min–Max	1–9	1–9	1–8
Skewness	–0.12	–0.23	–0.33
Kurtosis	–0.37	–0.12	–0.50
Activation level ¹			
<i>M (SD)</i>	5.16 (1.63)	5.52 (1.54)	5.29 (1.49)
Min–Max	1–9	1–9	1–9
Skewness	–0.42	–0.42	–0.38
Kurtosis	0.33	0.39	0.14
Activation level ²			
<i>M (SD)</i>	5.04 (1.73)	5.37 (1.56)	5.33 (1.39)
Min–Max	1–9	1–9	1–8
Skewness	–0.17	–0.46	–0.77
Kurtosis	–0.32	0.11	0.73
Activation level ³			
<i>M (SD)</i>	4.82 (1.70)	5.10 (1.67)	5.11 (1.46)
Min–Max	1–8	1–9	1–9
Skewness	–0.12	–0.32	–0.42
Kurtosis	–0.46	–0.19	0.40
Activation level ⁴			
<i>M (SD)</i>	4.94 (1.85)	5.19 (1.66)	4.93 (1.44)
Min–Max	1–9	1–9	1–8
Skewness	–0.07	–0.39	–0.37
Kurtosis	–0.58	–0.21	–0.13
Activation level ⁵			
<i>M (SD)</i>	4.78 (1.74)	5.24 (1.75)	5.14 (1.46)
Min–Max	1–9	1–9	1–9
Skewness	–0.10	–0.45	–0.19
Kurtosis	–0.62	0.01	–0.25
Activation level ^M			
<i>M (SD)</i>	4.95 (1.50)	5.29 (1.48)	5.16 (1.22)

Min–Max	1–8.80	1.20–9	1–7.80
Skewness	–0.04	–0.35	–0.44
Kurtosis	–0.10	0.21	0.40
Activation level change score			
<i>M</i> (<i>SD</i>)	–0.15 (1.60)	–0.10 (1.64)	0.05 (1.21)
Min–Max	–4.40–4	–5.20–4.20	–4–3
Skewness	–0.30	–0.17	–0.36
Kurtosis	0.41	0.73	1.08
Valence baseline			
<i>M</i> (<i>SD</i>)	5.58 (1.51)	5.97 (1.58)	5.80 (1.33)
Min–Max	1–9	2–9	1–8
Skewness	0.00	–0.35	–0.88
Kurtosis	–0.13	–0.50	1.23
Valence ¹			
<i>M</i> (<i>SD</i>)	5.43 (1.63)	6.00 (1.43)	5.92 (1.39)
Min–Max	1–9	1–9	1–8
Skewness	–0.04	–0.88	–0.87
Kurtosis	–0.06	1.64	0.81
Valence ²			
<i>M</i> (<i>SD</i>)	5.19 (1.63)	5.87 (1.62)	5.58 (1.39)
Min–Max	1–9	1–9	1–8
Skewness	0.13	–0.81	–0.88
Kurtosis	–0.29	0.95	1.31
Valence ³			
<i>M</i> (<i>SD</i>)	5.19 (1.65)	5.84 (1.56)	5.52 (1.40)
Min–Max	1–9	1–9	1–8
Skewness	0.13	–0.68	–0.75
Kurtosis	0.15	0.92	1.27
Valence ⁴			
<i>M</i> (<i>SD</i>)	5.40 (1.65)	5.70 (1.59)	5.48 (1.30)
Min–Max	1–9	1–9	2–8
Skewness	–0.15	–0.60	–0.34
Kurtosis	–0.14	0.77	–0.30
Valence ⁵			
<i>M</i> (<i>SD</i>)	5.25 (1.71)	5.67 (1.72)	5.54 (1.55)

Min–Max	1–9	1–9	1–9
Skewness	–0.35	–0.83	–0.31
Kurtosis	0.11	0.93	0.24
Valence ^M			
<i>M</i> (<i>SD</i>)	5.29 (1.49)	5.82 (1.39)	5.61 (1.12)
Min–Max	1.60–9	1.40–9	1.20–8
Skewness	0.13	–0.79	–0.53
Kurtosis	0.10	1.24	1.16
Valence change score			
<i>M</i> (<i>SD</i>)	–0.29 (1.14)	–0.16 (1.61)	–0.19 (1.15)
Min–Max	–5.20–2.60	–4.20–4.40	–3–4
Skewness	–1.04	0.41	0.20
Kurtosis	3.12	0.47	1.23

Note. ^M – average of the five responses after watching each video; change score – baseline measure subtracted from the average score

Similarly as before, paired samples *t*-tests were conducted independently in each group to assess variances from the baseline in various measures. In all groups, the majority of variables deviated from the assumption of normality.

Table 88: Baseline and post-intervention differences in the activation level and valence measures for the three experimental groups separately

	<i>t</i>	<i>p</i>	<i>Mean difference</i>	<i>95% CI</i>	<i>d</i>	<i>95% CI</i>
No music group						
Activation level ¹	–0.40	0.691	–0.06	–0.35–0.23	–0.04	–0.23–0.16
Activation level ²	0.37	0.712	0.06	–0.26–0.37	0.04	–0.16–0.23
Activation level ³	1.50	0.137	0.28	–0.09–0.64	0.15	–0.05–0.34
Activation level ⁴	0.75	0.455	0.16	–0.26–0.57	0.07	–0.12–0.27
Activation level ⁵	1.63	0.107	0.31	–0.07–0.70	0.16	–0.04–0.36
Activation level ^M	0.94	0.348	0.15	–0.17–0.46	0.09	–0.10–0.29
Valence ¹	1.31	0.195	0.15	–0.08–0.37	0.13	–0.07–0.32
Valence ²	2.79	0.006	0.39	0.11–0.67	0.28	0.08–0.47
Valence ³	3.07	0.003	0.39	0.14–0.65	0.30	0.11–0.50

Valence ⁴	1.28	0.205	0.18	−0.10–0.45	0.13	−0.07–0.32
Valence ⁵	2.24	0.027	0.32	0.04–0.61	0.22	0.03–0.12
Valence ^M	2.53	0.013	0.29	0.06–0.51	0.25	0.05–0.45
Calm music group						
Activation level ¹	−1.02	0.309	−0.13	−0.39–0.13	−0.10	−0.29–0.09
Activation level ²	0.12	0.906	0.02	−0.30–0.34	0.01	−0.18–0.20
Activation level ³	1.49	0.140	0.29	−0.10–0.67	0.15	−0.05–0.34
Activation level ⁴	1.07	0.288	0.20	−0.17–0.57	0.10	−0.09–0.30
Activation level ⁵	0.80	0.425	0.15	−0.23–0.53	0.08	−0.11–0.27
Activation level ^M	0.66	0.513	0.11	−0.21–0.42	0.06	−0.13–0.26
Valence ¹	−0.18	0.857	−0.03	−0.34–0.29	−0.02	−0.21–0.17
Valence ²	0.62	0.539	0.11	−0.23–0.44	0.06	−0.13–0.25
Valence ³	0.76	0.449	0.13	−0.22–0.48	0.07	−0.12–0.27
Valence ⁴	1.53	0.128	0.27	−0.08–0.61	0.15	−0.04–0.34
Valence ⁵	1.60	0.113	0.31	−0.07–0.68	0.16	−0.04–0.35
Valence ^M	1.00	0.322	0.16	−0.16–0.47	0.10	−0.10–0.29
Lively music group						
Activation level ¹	−1.50	0.137	−0.18	−0.42–0.06	−0.15	−0.35–0.05
Activation level ²	−1.48	0.142	−0.22	−0.52–0.08	−0.15	−0.35–0.05
Activation level ³	0.00	1.000	0.00	−0.29–0.29	0.00	−0.20–0.20
Activation level ⁴	1.23	0.222	0.18	−0.11–0.47	0.12	−0.07–0.32
Activation level ⁵	−0.20	0.843	−0.03	−0.33–0.27	−0.02	−0.22–0.18
Activation level ^M	−0.42	0.679	−0.05	−0.29–0.19	−0.04	−0.24–0.16
Valence ¹	−0.98	0.330	−0.12	−0.36–0.12	−0.10	−0.29–0.10
Valence ²	1.47	0.144	0.22	−0.08–0.52	0.15	−0.05–0.34
Valence ³	2.14	0.035	0.28	0.02–0.54	0.21	0.02–0.41
Valence ⁴	2.24	0.027	0.32	0.04–0.60	0.22	0.03–0.42
Valence ⁵	1.59	0.115	0.26	−0.07–0.59	0.16	−0.04–0.36
Valence ^M	1.67	0.097	0.19	−0.04–0.42	0.17	−0.03–0.36

Note. *CI* – confidence interval, *d* – effect size, $df_{No\ music} = 101$, $df_{Calm\ music} = 104$, $df_{Lively\ music} = 99$; ^M – average

Results in Table 88 reveal that there were no significant differences in activation level and valence for those who watched learning videos with calm background music and that there have been only marginal differences in valence scores within the control group and the group with lively music. In the control group, there was a significant decrease from baseline ($M = 5.58$, $SD = 1.51$) in the valence score after watching the second ($M = 5.19$, $SD = 1.63$), third ($M = 5.19$, $SD = 1.65$), and fifth video ($M = 5.26$, $SD = 1.71$), which was reflected also in the average difference from the valence baseline ($M = 5.29$, $SD = 1.49$). In the case of the lively music group, the difference from baseline ($M = 5.80$, $SD = 1.33$) was notable after the third ($M = 5.52$, $SD = 1.40$) and fourth video ($M = 5.48$, $SD = 1.30$), together with the average valence score ($M = 5.61$, $SD = 1.12$).

Next, differences in affective states between groups were assessed with ANCOVAs. Again, baseline measures were used as covariates, while the raw and average measurements were used as dependable variables (Table 89). In Appendix 20, the results of assumption checks can be seen. While in the case of the "Activation level 4" variable the homogeneity of variances assumption was not met, we still conducted ANCOVAs for all variables for simplicity's sake, while staying cautious when interpreting the variable in question.

Table 89: ANCOVA comparisons on activation level and valence items

	F	p	η^2p
Activation level ¹	0.682	0.506	0.004
Activation level ²	1.187	0.307	0.008
Activation level ³	0.932	0.395	0.006
Activation level ⁴	0.317	0.729	0.002
Activation level ⁵	1.734	0.178	0.011
Activation level ^M	1.007	0.367	0.007
Valence ¹	2.807	0.062	0.018
Valence ²	3.232	0.041	0.021
Valence ³	2.842	0.060	0.018
Valence ⁴	0.319	0.727	0.002
Valence ⁵	0.636	0.530	0.004
Valence ^M	2.172	0.116	0.014

Note. $df_1 = 2$, $df_2 = 303$; ^M – average

As there have been marginally significant differences in the case of three valence variables, post-hoc comparisons were conducted, which are represented in Table 90. As can be deducted, after applying the Bonferroni correction, the only significant differences are the differences in variables Valence 2 and 3 between the control group and the group watching videos with calm background music, as the latter reported higher valence compared to those who learned from videos without background music.

Table 90: Post-hoc comparisons for variables Valence 1, 2, and 3

Experimental groups		<i>t</i>	<i>p</i>	<i>p_{Bonferroni}</i>	Mean difference	<i>d</i>	95% <i>CI</i>
Valence ¹							
No music	Calm	-2.00	0.047	0.140	-0.34	-0.28	-0.56—0.01
No music	Lively	-2.11	0.036	0.108	-0.36	-0.30	-0.58—0.02
Calm	Lively	-0.13	0.900	1.000	-0.02	-0.02	-0.29—0.26
Valence ²							
No music	Calm	-2.53	0.012	0.035	-0.49	-0.35	-0.63—0.08
No music	Lively	-1.47	0.143	0.429	-0.28	-0.21	-0.49—0.07
Calm	Lively	1.05	0.294	0.881	0.20	0.15	-0.13—0.42
Valence ³							
No music	Calm	-2.38	0.018	0.053	-0.44	-0.33	-0.61—0.06
No music	Lively	-1.15	0.250	0.751	-0.22	-0.16	-0.44—0.12
Calm	Lively	1.22	0.223	0.668	0.23	0.17	-0.11—0.45

Note. *df* = 303

A MANCOVA, accompanied by a second set of ANCOVAs, was conducted by including potential confounding variables as covariates (prior interest, assessed prior knowledge, evaluated English proficiency, personality traits, and baseline measures of activation level and valence instead of the typical PANAVA-KS subscales). In the MANCOVA, the dependent variables comprised all five measurements of activation level and valence, excluding the general measures. Again, assumptions of normality and homogeneity were not met ($\chi^2(110) = 200.95$, $p < 0.001$, $W = 0.76$, $p < 0.001$), but a MANCOVA was conducted due to the large sample size. The analysis uncovered a marginally significant effect (Wilks' Lambda = 0.87, $F(20, 554) = 2.08$, $p = 0.004$).

Table 91: ANCOVA/Quade comparisons (with multiple covariates) on activation level and valence items

	ANCOVA*/Quade's test**			Homogeneity test***		Normality test	
	<i>F</i>	<i>p</i>	η^2p	<i>F</i>	<i>p</i>	<i>W</i>	<i>p</i>
Activation level ¹	0.89	0.411	0.01	0.50	0.605	0.98	< .001
Activation level ²	0.93	0.395	0.01	1.55	0.214	0.98	< .001
Activation level ^{3**}	1.02	0.363		3.53	0.031	0.99	0.006
Activation level ⁴	0.13	0.881	0.00	2.94	0.055	0.99	0.005
Activation level ⁵	1.96	0.143	0.01	2.59	0.077	0.99	0.003
Activation level ^{M**}	1.01	0.367		4.38	0.013	0.98	0.013
Valence ¹	3.88	0.022	0.03	0.41	0.666	0.94	< .001
Valence ²	2.48	0.085	0.02	2.20	0.112	0.97	< .001
Valence ^{3**}	4.04	0.019		3.95	0.020	0.98	0.001
Valence ^{4**}	0.24	0.789		5.47	0.005	0.99	0.004
Valence ⁵	0.99	0.373	0.01	1.99	0.138	0.96	< .001
Valence ^{M**}	3.47	0.032		6.17	0.002	0.98	< .001

Note. * $df_1 = 2$, $df_2 = 286$; ** $df_1 = 2$, $df_2 = 296$; *** $df_1 = 2$, $df_2 = 296$; ^M – average

Table 91 presents the results of multiple ANCOVAs along with assumption checks. For five variables, neither assumption was met, so Quade's non-parametric tests were performed instead. As can be seen, there were no significant differences in most of the variables. However, there were some marginal differences between the first three valence ratings and the average one, so post-hoc comparisons were made. Table 92 shows that there were (marginally) significant differences between the no music and calm music conditions in all those variables and a significant difference between the no music and lively music conditions in Valence 1.

Table 92: ANCOVA post-hoc comparisons for Valence 1, 2, 3, and average

Experimental groups		<i>t</i>	<i>p</i>	<i>p</i> _{Bonferroni}	Mean difference	<i>d</i>	95% <i>CI</i>
Valence ^{1*}							
No music	Calm	-2.29	0.023	0.069	-0.39	-0.33	-0.61—0.04
No music	Lively	-2.51	0.013	0.038	-0.43	-0.36	-0.64—0.08
Calm	Lively	-0.22	0.823	1.000	-0.04	-0.03	-0.32—0.25
Valence ^{2*}							
No music	Calm	-2.20	0.029	0.086	-0.41	-0.31	-0.60—0.03
No music	Lively	-1.41	0.160	0.479	-0.26	-0.20	-0.48—0.08
Calm	Lively	0.78	0.436	1.000	0.15	0.11	-0.17—0.40
Valence ^{3**}							
No music	Calm	-2.84	0.005				
No music	Lively	-1.39	0.166				
Calm	Lively	1.45	0.149				
Valence ^{M**}							
No music	Calm	-2.62	0.009				
No music	Lively	-1.51	0.133				
Calm	Lively	1.11	0.267				

Note. **df* = 286, ***df* = 296; ^M – average

All in all, our results replicate the mixed and complicated nature of findings from previous studies (e.g., Du et al., 2020; Jäncke and Sandmann, 2010; Lehmann and Seufert, 2017). While it mostly seems that both types of music failed to elicit significant emotional responses from the participants (measured with two one item questions regarding valence and activation level), in some cases, results approached significance. The lack of significant results differs from our pre-study findings, where participants listening to the calm and lively songs felt differences in their activation levels but not in pleasantness. This suggests that while music alone can significantly influence students' emotional states, its impact diminishes when used as background music with the focus on learning content.

Interest in the topic

Interest in the topic was measured in two ways: through a brief questionnaire immediately after watching the videos and a question at the beginning of the delayed

post-test, which 38.44% of participants completed after a week. Table 93 presents descriptive statistics for the two variables.

Table 93: Descriptive statistics of the two interest variables divided by group

	No music	Calm music	Lively music
Situational interest			
<i>M (SD)</i>	4.08 (1.25)	4.46 (1.28)	4.21 (1.24)
Min–Max	1–6.67	1–6.33	1–7
Skewness	–0.04	–0.70	0.04
Kurtosis	–0.74	–0.08	–0.13
Delayed interest			
<i>M (SD)</i>	4.30 (1.62)	4.46 (1.72)	4.49 (1.49)
Min–Max	1–7	1–7	1–7
Skewness	–0.71	–0.58	–0.55
Kurtosis	–0.52	–0.79	–0.18

Appendix 20 reveals that while the variables did not meet the assumption of normality, they did not violate the homogeneity assumption, so we proceeded with ANOVAs. Comparisons indicated a marginally significant difference among the three groups in situational interest (measured after watching the videos; $F(2,304) = 2.379$, $p = 0.094$, $n^2p = 0.015$), but not in the interest that was assessed a week after watching the videos ($F(2,115) = 0.160$, $p = 0.852$, $n^2p = 0.003$).

Table 94: Post-hoc comparisons for Situational interest

Experimental groups		<i>t</i>	<i>p</i>	<i>p</i> _{Bonferroni}	Mean difference	<i>d</i>	95% <i>CI</i>
No music	Calm	–2.15	0.032	0.097	–0.38	–0.30	–0.57—0.02
No music	Lively	–0.75	0.452	1.000	–0.13	–0.11	–0.38—0.17
Calm	Lively	1.38	0.169	0.506	0.24	0.19	–0.08—0.47

Note. *df* = 304

Post-hoc comparisons (Table 94) reveal a marginally significant higher situational interest in the group who watched videos with added calm background music compared to the group who watched videos without any music added. This difference, however, was only noticeable immediately after the learning session.

ANCOVAs reveal a similar image, with Situational interest showing a marginally significant effect ($F(2, 285) = 3.10, p = 0.047, n^2p = 0.02; W = 0.99, p = 0.270; F(1, 296) = 0.37, p = 0.691$), which was not the case for delayed interest ($F(2,104) = 1.81, p = 0.169, n^2p = 0.03; W = 0.98, p = 0.119; F(1, 115) = 2.71, p = 0.071$). Post-hoc tests for the first variable (Table 95) again showed the difference between the no music and calm music conditions, indicating that the presence of calm background music may create a more favorable and relaxing learning environment, reducing negative activating emotions such as stress or anxiety,

allowing students to immerse themselves more into the topic. While this result is marginally significant and speculative, it warrants further research.

Table 95: ANCOVA post-hoc comparisons for Situational interest

Experimental groups		<i>t</i>	<i>p</i>	<i>p</i> _{Bonferroni}	Mean difference	<i>d</i>	95% <i>CI</i>
No music	Calm	-2.34	0.020	0.059	-0.32	-0.33	-0.62—0.05
No music	Lively	-0.42	0.672	1.000	-0.06	-0.06	-0.34—0.22
Calm	Lively	1.89	0.060	0.180	0.26	0.27	-0.01—0.56

Note. *df* = 285

Intrinsic motivation

Background music was also predicted to affect participants' motivation. An ANOVA revealed no significant results in motivation levels ($F(2,304) = 1.40, p = 0.249, n^2p = 0.01$), as participants from the control group ($M = 4.17, SD = 1.31$) and the groups with calm ($M = 4.47, SD = 1.27$) and lively background music ($M = 4.36, SD = 1.32$) did not significantly differ in their motivation levels. The result was the same even after including eleven variables as covariates ($F(2,285) = 1.39, p = 0.252, n^2p = 0.01; W = 0.99, p = 0.003; F(1, 296) = 1.31, p = 0.272$).

Learners' experience

To gauge learners' experiences with the learning videos, a set of five questions commonly employed in multimedia learning studies was used. These questions assessed participants' motivation to pay attention, perceived difficulty of the lectures, expended effort in learning, enjoyment of the experience, and their interest in having more lessons similar to the one they just viewed. The correlation matrix in Appendix 19 illustrates the relationships between these variables, with some exhibiting marginally

significant correlation (e.g., exerting more effort and enjoyment, $r_{\text{effort} - \text{enjoyment}} = 0.110$, $p = 0.055$) and others demonstrating strong correlations (e.g., $r_{\text{enjoyment} - \text{paying attention}} = 0.728$, $p < .001$; $r_{\text{enjoyment} - \text{more lessons}} = 0.768$, $p < .001$). Descriptive statistics for these questions are provided in Table 96.

Table 96: Descriptive statistics of the learners' experience variables divided by group

	No music	Calm music	Lively music
Paying attention			
<i>M (SD)</i>	3.95 (1.61)	4.50 (1.56)	4.24 (1.39)
Min–Max	1–7	1–7	1–7
Skewness	–0.02	–0.45	–0.24
Kurtosis	–1.12	–0.53	–0.81
Difficulty			
<i>M (SD)</i>	2.86 (1.44)	2.79 (1.34)	2.80 (1.15)
Min–Max	1–7	1–7	1–6
Skewness	0.73	0.67	0.84
Kurtosis	0.10	–0.01	0.25
Exerting more effort			
<i>M (SD)</i>	3.42 (1.63)	3.55 (1.51)	3.54 (1.30)
Min–Max	1–7	1–7	1–6
Skewness	0.24	0.20	–0.19
Kurtosis	–0.97	–0.46	–0.98
Enjoyment			
<i>M (SD)</i>	4.30 (1.57)	4.70 (1.48)	4.52 (1.34)
Min–Max	1–7	1–7	1–7
Skewness	–0.19	–0.65	–0.41
Kurtosis	–0.99	–0.17	–0.21
More lessons like this			
<i>M (SD)</i>	4.05 (1.84)	4.62 (1.55)	4.32 (1.52)
Min–Max	1–7	1–7	1–7
Skewness	–0.08	–0.72	–0.06
Kurtosis	–1.08	–0.18	–0.70

Assumption checks (Appendix 20) revealed that three of the five variables had unequal variances. In these cases, Welch's ANOVA was performed instead of Fisher's, as it does not assume equal variances. All results are presented in Table 97.

Table 97: Comparisons of the three groups on learners' experience variables

	<i>F</i>	<i>df</i> ₁	<i>df</i> ₂	<i>p</i>	η^2p
Paying attention	3.31	2	304	0.038	0.02
Difficulty*	0.08	2	201.39	0.922	0.00
Exerting more effort*	0.22	2	201.31	0.805	0.00
Enjoyment	1.84	2	304	0.160	0.01
More lessons like this*	2.96	2	201.18	0.054	0.02

Note. * Welch's ANOVA test results

Given the marginally significant differences in the variables "Paying attention" and "More lessons like this," post-hoc tests were conducted (see Table 98 for results). A Tukey post-hoc test was used for the "Paying attention" variable, assuming equal variances, while a Games-Howell post-hoc test was conducted for the "More lessons like this" variable, considering unequal variances.

Table 98: Post-hoc comparisons for the Paying attention and More lessons like this variables

Experimental groups		<i>t</i>	<i>df</i>	<i>p</i> _{Bonferroni}	Mean difference	<i>d</i>	95% <i>CI</i>
Paying attention							
No music	Calm	-2.57	304	0.032	-0.54	-0.36	-0.63—0.08
No music	Lively	-1.35	304	0.536	-0.29	-0.19	-0.47—0.09
Calm	Lively	1.20	304	0.694	0.26	0.17	-0.11—0.44
More lessons like this							
No music	Calm	-2.41	197.19	0.039	-0.57	-0.35	-0.62—0.07
No music	Lively	-1.14	194.41	0.723	-0.27	-0.17	-0.44—0.11
Calm	Lively	1.40	202.82	0.578	0.30	0.18	-0.09—0.46

Note. Tukey post-hoc test was performed on the Paying attention variable and Games-Howell post-hoc test was conducted for the *More lessons like this* variable

In both cases, post-hoc comparisons revealed no significant differences between the control group and the lively music group and the calm and lively music group. However, participants in the calm music group reported paying significantly more attention to the videos and expressed a higher desire for more lessons like the ones they just had compared to the control group with no background music. This is consistent with previous analysis, which showed that the calm music group also expressed higher interest in the videos compared to the control group.

Additionally, a MANCOVA was performed for all five variables together, to include the potential effect of confounding variables, which did not show a significant effect (Wilks' Lambda = 0.96, $F(5, 532) = 1.210$, $p = 0.272$; $\chi^2(30) = 47.37$, $p = 0.023$, $W = 0.97$, $p < 0.001$). Results of univariate tests (ANCOVAs) are presented in Table 99.

Table 99: ANCOVA comparisons with eleven covariates on the learners' experience variables

	ANCOVA*			Homogeneity test**		Normality test	
	<i>F</i>	<i>p</i>	η^2p	<i>F</i>	<i>p</i>	<i>W</i>	<i>p</i>
Paying attention	4.47	0.012	0.03	1.23	0.292	0.99	0.005
Difficulty	0.01	0.993	0.00	1.39	0.250	0.97	< .001
Exerting more effort	0.66	0.518	0.00	0.56	0.572	0.99	0.130
Enjoyment	2.66	0.072	0.02	1.99	0.139	0.99	0.022
More lessons like this	3.43	0.034	0.02	2.33	0.100	0.99	0.157

Note. * $df_1 = 2$, $df_2 = 285$; ** $df_1 = 2$, $df_2 = 296$

Three variables had marginally significant results, for which post-hoc tests were made (Table 100). As before, the only noticeable differences were between the control and calm music experimental groups. These findings further support the idea that calm music may help regulate unpleasant activating emotions, making it easier for students to focus on the lesson, enjoy it more, and increasing their interest in similar future lessons.

Table 100: ANCOVA post-hoc comparisons for the Paying attention, Enjoyment, and More lessons like this variables

Experimental groups		<i>t</i>	<i>p</i>	<i>p</i> _{Bonferroni}	Mean difference	<i>d</i>	95% <i>CI</i>
Paying attention							
No music	Calm	-2.99	0.003	0.009	-0.56	-0.43	-0.71—0.14
No music	Lively	-1.37	0.173	0.519	-0.26	-0.20	-0.48—0.09
Calm	Lively	1.59	0.114	0.342	0.30	0.23	-0.06—0.51
Enjoyment							
No music	Calm	-2.30	0.022	0.066	-0.39	-0.33	-0.61—0.05
No music	Lively	-1.25	0.213	0.639	-0.21	-0.18	-0.46—0.10
Calm	Lively	1.03	0.305	0.916	0.17	0.15	-0.14—0.43
More lessons like this							
No music	Calm	-2.62	0.009	0.028	-0.53	-0.37	-0.66—0.09
No music	Lively	-1.33	0.186	0.558	-0.27	-0.19	-0.47—0.09
Calm	Lively	1.26	0.208	0.623	0.26	0.18	-0.10—0.47

Note. *df* = 285

3.5.3.4 Cognitive outcomes

Variables assessing cognitive outcomes are divided into perceived cognitive load and mental effort. The correlational matrix in Appendix 19 demonstrates that the correlations between various types of cognitive load and the overall mental effort measure ranged from insignificant to moderate ($-0.211 < r < 0.414$).

Cognitive load

In line with the literature review, it was anticipated that there would be a significant difference in extraneous cognitive load levels between the groups with different narrators (Hypothesis 10). Descriptive statistics, categorized by type of cognitive load, are presented in Table 101.

Table 101: Descriptive statistics of the cognitive load questionnaire divided by group

	No music	Calm music	Lively music
Intrinsic cognitive load			
<i>M (SD)</i>	3.50 (1.27)	3.51 (1.38)	3.63 (1.16)
Min–Max	1–6.50	1–7	1–6
Skewness	0.40	0.16	0.09
Kurtosis	–0.61	–0.78	–0.64
Extraneous cognitive load			
<i>M (SD)</i>	3.06 (1.33)	2.87 (1.17)	3.03 (1.15)
Min–Max	1–6.67	1–6.67	1–6.33
Skewness	0.58	0.46	0.50
Kurtosis	–0.26	0.05	–0.10
Germane cognitive load			
<i>M (SD)</i>	4.73 (1.34)	4.99 (1.17)	4.86 (1.11)
Min–Max	1–7	1–7	1–7
Skewness	–0.34	–0.69	–0.86
Kurtosis	–0.41	0.49	1.42

Assumption checks (Appendix 20) indicated unequal variances in the case of germane cognitive load, so Welch’s ANOVA was performed as it does not assume equal variances (Table 102). No significant differences were found between the three groups regarding their cognitive load, leading us to reject Hypothesis 10. While one study reported that the impact of seductive details on cognitive load and learning might be moderated by arousal (Schneider et al., 2019), our results did not demonstrate any differences between the effects of calming and lively music.

Table 102: Comparisons of the three groups on learners’ experience variables

	<i>F</i>	<i>df</i> ₁	<i>df</i> ₂	<i>p</i>	<i>η</i> ² <i>p</i>
Intrinsic cognitive load	0.30	2	304	0.741	0.00
Extraneous cognitive load	0.76	2	304	0.468	0.01
Germane cognitive load*	1.11	2	201.50	0.331	0.01

Note. * Welch’s ANOVA test results

A MANCOVA with eleven covariates did not reveal a significant effect on the three cognitive load variables (Wilks' Lambda = 0.98, $F(6, 566) = 0.82$, $p = 0.556$; $\chi^2(12) = 16.82$, $p = 0.157$, $W = 0.98$, $p < .001$). Additional ANCOVAs were conducted, further failing to find any meaningful differences in intrinsic ($F(2,285) = 0.17$, $p = 0.842$, $n^2p = 0.00$; $W = 0.98$, $p = 0.003$; $F(1, 296) = 1.31$, $p = 0.273$), extraneous ($F(2,285) = 1.00$, $p = 0.370$, $n^2p = 0.01$; $W = 0.96$, $p < .001$; $F(1, 296) = 1.97$, $p = 0.141$), and germane cognitive load ($F(2,285) = 1.00$, $p = 0.371$, $n^2p = 0.01$; $W = 0.98$, $p < .001$; $F(1, 296) = 0.99$, $p = 0.372$).

Mental effort

Following the viewing of each video, participants also provided feedback on the mental effort invested in comprehending the learning content, resulting in five distinct mental effort measures. Descriptive statistics for these five measures, along with their average, are presented in Table 103.

Table 103: Descriptive statistics of the mental effort ratings divided by group

	No music	Calm music	Lively music
Mental effort ¹			
<i>M (SD)</i>	4.25 (1.71)	4.62 (1.72)	4.43 (1.62)
Min–Max	1–8	1–9	1–7
Skewness	0.08	–0.07	–0.32
Kurtosis	–0.49	–0.19	–0.86
Mental effort ²			
<i>M (SD)</i>	4.50 (1.69)	4.82 (1.68)	4.68 (1.54)
Min–Max	1–9	1–9	1–8
Skewness	0.23	–0.32	–0.47
Kurtosis	–0.34	0.04	–0.43
Mental effort ³			
<i>M (SD)</i>	4.41 (1.74)	4.72 (1.67)	4.50 (1.52)
Min–Max	1–9	1–9	1–8
Skewness	0.00	–0.34	–0.40
Kurtosis	–0.30	–0.03	–0.28
Mental effort ⁴			
<i>M (SD)</i>	4.22 (1.69)	4.64 (1.62)	4.37 (1.45)
Min–Max	1–9	1–9	1–8

Skewness	0.20	−0.30	−0.43
Kurtosis	0.03	0.22	0.13
Mental effort ⁵			
<i>M</i> (<i>SD</i>)	4.20 (1.59)	4.57 (1.75)	4.37 (1.54)
Min–Max	1–9	1–9	1–8
Skewness	0.17	−0.34	−0.29
Kurtosis	0.07	−0.33	−0.03
Mental effort ^{<i>M</i>}			
<i>M</i> (<i>SD</i>)	4.32 (1.48)	4.67 (1.54)	4.47 (1.36)
Min–Max	1.40–8.80	1–8	1–7.80
Skewness	0.37	−0.40	−0.41
Kurtosis	0.25	−0.01	−0.03

Note. ^{*M*} – average

As the assumption of homogeneity of variances was met for all variables, Fisher's ANOVAs were performed to compare the groups in their mental effort levels.

Table 104: ANOVA comparisons of the three groups on mental effort

	<i>F</i>	<i>p</i>	η^2p
Mental effort ¹	1.210	0.300	0.008
Mental effort ²	0.984	0.375	0.006
Mental effort ³	0.990	0.373	0.006
Mental effort ⁴	1.863	0.157	0.012
Mental effort ⁵	1.369	0.256	0.009
Mental effort ^{<i>M</i>}	1.566	0.211	0.010

Note. $df_1 = 2$, $df_2 = 304$; ^{*M*} – average

As was the case with the cognitive load questionnaire, there were no significant differences between the groups in perceived mental effort even when measured after each video (Table 104). No post-hoc comparisons were made.

A MANCOVA with five measures of mental effort as dependent variables and prior knowledge, interest, initial emotional state, English proficiency, and personality as covariates did not reveal any significant effect (Wilks' Lambda = 0.99, $F(10, 562) = 0.41$, $p = 0.943$; $\chi^2(30) = 42.83$, $p = 0.061$, $W = 0.95$, $p < .001$). These results were further confirmed with univariate ANCOVAs (Table 105).

Table 105: ANCOVA comparisons with eleven covariates on mental effort

	ANCOVA*			Homogeneity test**		Normality test	
	<i>F</i>	<i>p</i>	η^2p	<i>F</i>	<i>p</i>	<i>W</i>	<i>p</i>
Mental effort ¹	1.61	0.201	0.01	0.68	0.506	0.99	0.035
Mental effort ²	1.20	0.302	0.01	0.04	0.963	0.99	0.006
Mental effort ³	1.22	0.298	0.01	0.79	0.455	0.99	0.060
Mental effort ⁴	1.82	0.164	0.01	0.62	0.538	0.99	0.034
Mental effort ⁵	1.54	0.216	0.01	1.01	0.366	0.99	0.008
Mental effort ^M	1.85	0.160	0.01	1.13	0.324	0.99	0.064

Note. * $df_1 = 1$, $df_2 = 285$; ** $df_1 = 1$, $df_2 = 296$; ^M – average

3.5.3.5 Learning

As was done in the case of the previous experiment, the learning outcomes are divided into those that were assessed in the same testing session as the videos viewing and a week after the videos viewing. Learning outcomes include several objective and subjective measures, such as self-evaluated learning, self-evaluated test performance, and assessed knowledge, together with the level of confidence in each answer. The same variables (without self-evaluated learning) were assessed also seven days after the learning session, and the corresponding results are detailed in the subsection on the delayed experiment.

Objective and subjective test performance in the immediate part of the experiment

Table 106 presents detailed statistics for all learning-related outcomes during the main experiment. The "knowledge" variable encompasses the total points obtained on the test, while the "retention" and "transfer" variables comprise points acquired for correctly answering questions related to retention and transfer, respectively.

There was a low to moderate correlation between self-evaluated learning from the videos with all other learning variables ($0.205 < r < 0.491$, $p < .001$) and a low to moderate correlation between participants' self-evaluated test performance with actual scores on the retention ($r = 0.343$, $p < .001$) and transfer ($r = 0.290$, $p < .001$) segments of the test (see Appendix 19).

Table 106: Descriptive statistics of the learning outcomes from the immediate part of the experiment divided by group

	No music	Calm music	Lively music
Self-evaluated learning			
<i>M (SD)</i>	4.11 (1.35)	4.50 (1.08)	4.31 (1.08)
Min–Max	1–7	1–7	1–7
Skewness	–0.18	–0.20	0.04
Kurtosis	0.22	1.56	0.54
Knowledge			
<i>M (SD)</i>	15.63 (4.32)	15.81 (5.12)	16.22 (5.02)
Min–Max	3–25	0–26	6–28
Skewness	–0.22	–0.30	0.18
Kurtosis	–0.17	0.02	–0.57
Retention			
<i>M (SD)</i>	10.22 (3.10)	10.23 (3.97)	10.52 (3.67)
Min–Max	3–16	0–18	4–18
Skewness	–0.18	–0.21	0.37
Kurtosis	–0.63	–0.22	–0.78
Transfer			
<i>M (SD)</i>	5.47 (1.83)	5.48 (1.76)	5.71 (1.89)
Min–Max	0–10	0–9	1–10
Skewness	–0.35	–0.37	–0.20
Kurtosis	0.32	0.12	–0.19
Certainty in all answers			
<i>M (SD)</i>	61.37 (18.80)	65.04 (20.63)	67.51 (17.62)
Min–Max	9.14–96.55	0–99.14	6.90–97.41
Skewness	–0.54	–1.08	–0.72
Kurtosis	–0.30	0.94	0.87
Certainty in correct answers			
<i>M (SD)</i>	63.82 (19.82)	68.78 (21.52)	70.44 (18.39)
Min–Max	9–99.76	0–100	6.25–97.22
Skewness	–0.55	–1.17	–1.05
Kurtosis	–0.37	1.10	1.28
Certainty in incorrect answers			
<i>M (SD)</i>	57.01 (18.50)	59.88 (19.71)	62.97 (17.16)

Min–Max	9.29–90.91	0–100	7.69–100
Skewness	–0.46	–0.88	–0.33
Kurtosis	–0.34	0.81	0.43
Certainty in all retention answers			
<i>M (SD)</i>	61.21 (19.55)	64.69 (20.40)	66.65 (17.92)
Min–Max	9.47–94.74	0–98.68	5.26–96.05
Skewness	–0.53	–1.00	–0.58
Kurtosis	–0.46	0.77	0.71
Certainty in correct retention answers			
<i>M (SD)</i>	63.77 (20.95)	68.55 (21.62)	70.54 (18.50)
Min–Max	8.89–99.62	0–100	5.26–97.33
Skewness	–0.56	–1.06	–0.98
Kurtosis	–0.29	0.76	1.00
Certainty in incorrect retention answers			
<i>M (SD)</i>	56.34 (19.07)	58.52 (19.43)	61.00 (18.47)
Min–Max	10–90	0–100	0–100
Skewness	–0.36	–0.71	–0.23
Kurtosis	–0.61	0.59	0.38
Certainty in all transfer answers			
<i>M (SD)</i>	62.00 (20.59)	65.70 (22.89)	69.13 (19.15)
Min–Max	2.80–100	0–100	10–100
Skewness	–0.66	–0.95	–0.94
Kurtosis	–0.12	0.64	0.64
Certainty in correct transfer answers			
<i>M (SD)</i>	64.65 (22.53)	68.47 (24.62)	71.52 (20.57)
Min–Max	4–100	0–100	0–100
Skewness	–0.59	–0.94	–1.13
Kurtosis	–0.46	0.33	1.26
Certainty in incorrect transfer answers			
<i>M (SD)</i>	58.26 (21.29)	62.38 (23.49)	65.60 (19.96)
Min–Max	2.29–100	0–100	17.29–100
Skewness	–0.38	–0.56	–0.48
Kurtosis	–0.17	0.10	–0.37
Self-evaluated test performance			
<i>M (SD)</i>	3.60 (1.17)	3.87 (1.35)	3.89 (1.19)

Min–Max	1–7	1–7	1–7
Skewness	–0.22	0.01	–0.00
Kurtosis	0.36	0.04	0.55

While the variables in question mostly did not meet the assumption of normality, the assumption of homogeneity was not violated by either of them (Appendix 20), so ANOVAs were performed to compare the groups (Table 107).

Table 107: ANOVA comparisons of the three groups on learning outcomes in the immediate part of the experiment

	<i>F</i>	<i>df</i> ₂	<i>p</i>	η^2p
Self-evaluated learning	2.940	304	0.054	0.019
Knowledge	0.398	300	0.672	0.003
Retention	0.218	301	0.804	0.001
Transfer	0.549	299	0.578	0.004
Certainty	2.626	298	0.074	0.017
Certainty in correct answers	2.984	298	0.052	0.020
Certainty in incorrect answers	2.597	298	0.076	0.017
R Certainty	2.046	299	0.131	0.013
R Certainty in correct answers	2.933	299	0.055	0.019
R Certainty in incorrect answers	1.518	299	0.221	0.010
T Certainty	2.900	298	0.057	0.019
T Certainty in correct answers	2.305	297	0.102	0.015
T Certainty in incorrect answers	2.853	295	0.059	0.019
Self-evaluated test performance	1.671	298	0.190	0.011

Note. *df*₁ = 2; R – retention, T – transfer

While there were no significant differences in the actual test scores and self-evaluated test performance, there were some marginally significant differences in some of the subjective learning outcomes, such as participants' perception of how much they learned from the videos and their certainty in their answers. For these variables, post-hoc comparisons were made (Table 108).

Table 108: Post-hoc comparisons for learning outcomes in the immediate part of the experiment

Experimental groups		<i>t</i>	<i>p</i>	<i>p</i> _{Bonferroni}	Mean difference	<i>d</i>	95% <i>CI</i>
Self-evaluated learning (<i>df</i> ₂ = 304)							
No music	Calm	-2.43	0.016	0.048	-0.40	-0.34	-0.61—0.06
No music	Lively	-1.22	0.223	0.670	-0.20	-0.17	-0.45—0.11
Calm	Lively	1.18	0.237	0.712	0.20	0.17	-0.11—0.44
Certainty (<i>df</i> ₂ = 298)							
No music	Calm	-1.37	0.173	0.518	-3.67	-0.19	-0.47—0.08
No music	Lively	-2.28	0.024	0.071	-6.14	-0.32	-0.60—0.04
Calm	Lively	-0.92	0.361	1.000	-2.47	-0.13	-0.41—0.15
Certainty in correct answers (<i>df</i> ₂ = 298)							
No music	Calm	-1.76	0.079	0.236	-4.96	-0.25	-0.53—0.03
No music	Lively	-2.34	0.020	0.059	-6.62	-0.33	-0.61—0.05
Calm	Lively	-0.59	0.557	1.000	-1.66	-0.08	-0.36—0.20
Certainty in incorrect answers (<i>df</i> ₂ = 298)							
No music	Calm	-1.10	0.271	0.814	-2.87	-0.16	-0.43—0.12
No music	Lively	-2.28	0.023	0.070	-5.96	-0.32	-0.60—0.04
Calm	Lively	-1.18	0.238	0.714	-3.09	-0.17	-0.45—0.11
R Certainty in correct answers (<i>df</i> ₂ = 299)							
No music	Calm	-1.67	0.096	0.289	-4.78	-0.23	-0.51—0.04
No music	Lively	-2.35	0.019	0.058	-6.77	-0.33	-0.61—0.05
Calm	Lively	-0.69	0.490	1.000	-1.99	-0.10	-0.38—0.18
T Certainty (<i>df</i> ₂ = 298)							
No music	Calm	-1.26	0.210	0.630	-3.70	-0.18	-0.45—0.10
No music	Lively	-2.41	0.017	0.050	-7.13	-0.34	-0.62—0.06
Calm	Lively	-1.16	0.248	0.743	-3.43	-0.16	-0.44—0.11
T Certainty in correct answers (<i>df</i> ₂ = 297)							
No music	Calm	-1.20	0.232	0.696	-3.83	-0.17	-0.45—0.11
No music	Lively	-2.14	0.033	0.099	-6.88	-0.30	-0.58—0.02
Calm	Lively	-0.95	0.342	1.000	-3.05	-0.13	-0.41—0.14
T Certainty in incorrect answers (<i>df</i> ₂ = 295)							
No music	Calm	-1.35	0.178	0.534	-4.12	-0.19	-0.47—0.09
No music	Lively	-2.38	0.018	0.054	-7.35	-0.34	-0.62—0.06

Calm	Lively	-1.05	0.296	0.888	-3.22	-0.15	-0.43-0.13
------	--------	-------	-------	-------	-------	-------	------------

Note. $df_1 = 2$; R – retention, T – transfer

Participants who watched videos with calm background music differed significantly from those without background music, as the former group perceived that they gained significantly more knowledge than the control group. Conversely, the lively music group showed a marginal increase in certainty regarding their answers, whether correct or incorrect, compared to the group with no music.

Following the pattern from before, additional tests were made to assess results while accounting for eleven confounding variables. Initially, a MANCOVA was performed with the same eleven covariates as before and the following (six) dependant variables: outcomes for the retention and transfer sections of the test, levels of certainty in correct and incorrect answers (segmented by retention and transfer), self-evaluated learning, and self-evaluated test performance. No statistically significant effect was observed (Wilks' Lambda = 0.95, $F(12, 560) = 1.29$, $p = 0.218$; $\chi^2(42) = 67.03$, $p = 0.008$, $W = 0.98$, $p < .001$).

Table 109: ANCOVA comparisons with eleven covariates on learning variables in the immediate part of the experiment

	ANCOVA*			Homogeneity test**		Normality test	
	<i>F</i>	<i>p</i>	η^2p	<i>F</i>	<i>p</i>	<i>W</i>	<i>p</i>
Knowledge	0.24	0.783	0.00	0.91	0.405	0.99	0.204
Retention	0.10	0.903	0.00	1.60	0.203	1.00	0.530
Transfer	0.37	0.691	0.00	1.47	0.232	0.99	0.051
Certainty	2.53	0.081	0.02	2.43	0.090	0.98	< .001
Certainty in correct answers	2.71	0.068	0.02	2.16	0.118	0.97	< .001
Certainty in incorrect answers	2.55	0.080	0.02	1.34	0.264	0.99	0.003
R Certainty	1.99	0.139	0.01	1.27	0.283	0.99	0.008
R Certainty in correct answers	2.80	0.062	0.02	1.03	0.357	0.98	< .001
R Certainty in incorrect answers	1.45	0.236	0.01	0.42	0.659	0.99	0.078
T Certainty	2.78	0.064	0.02	1.49	0.228	0.97	< .001

T Certainty in correct answers	2.22	0.111	0.02	0.97	0.380	0.98	< .001
T Certainty in incorrect answers	3.02	0.050	0.02	2.07	0.128	0.99	0.017
Self-evaluated learning	3.22	0.041	0.02	2.09	0.126	0.98	0.002
Self-evaluated test performance	1.13	0.325	0.01	0.88	0.417	0.99	0.300

Note. * $df_1 = 2$, $df_2 = 285$; ** $df_1 = 2$, $df_2 = 296$; R – retention, T – transfer

Subsequent univariate ANCOVAs were conducted for all learning outcome variables, as detailed in Table 109. While no main effects were found at the Bonferroni level, there were some (subjective) variables that indicated a possible marginally significant effect. For those variables, post-hoc comparisons were made (Table 110).

Table 110: ANCOVA post-hoc comparisons for some of the learning outcome variables in the immediate part of the experiment

Experimental groups		<i>t</i>	<i>p</i>	<i>p</i> _{Bonferroni}	Mean difference	<i>d</i>	95% <i>CI</i>
Certainty							
No music	Calm	–1.10	0.270	0.811	–2.54	–0.16	–0.44–0.12
No music	Lively	–2.25	0.025	0.075	–5.22	–0.32	–0.61–0.04
Calm	Lively	–1.15	0.249	0.748	–2.69	–0.17	–0.45–0.12
Certainty in correct answers							
No music	Calm	–1.57	0.118	0.354	–3.81	–0.22	–0.50–0.06
No music	Lively	–2.27	0.024	0.072	–5.58	–0.33	–0.61–0.04
Calm	Lively	–0.72	0.474	1.000	–1.77	–0.10	–0.39–0.18
Certainty in incorrect answers							
No music	Calm	–0.86	0.393	1.000	–2.00	–0.12	–0.40–0.16
No music	Lively	–2.24	0.026	0.077	–5.30	–0.32	–0.61–0.04
Calm	Lively	–1.39	0.166	0.497	–3.30	–0.20	–0.49–0.08
R Certainty in correct answers							
No music	Calm	–1.54	0.124	0.371	–3.85	–0.22	–0.50–0.06
No music	Lively	–2.32	0.021	0.063	–5.84	–0.33	–0.62–0.05
Calm	Lively	–0.79	0.430	1.000	–2.00	–0.11	–0.40–0.17
T Certainty							

No music	Calm	−0.96	0.338	1.000	−2.54	−0.14	−0.42−0.14
No music	Lively	−2.35	0.020	0.059	−6.28	−0.34	−0.62—0.05
Calm	Lively	−1.39	0.165	0.495	−3.74	−0.20	−0.49−0.08
T Certainty in incorrect answers							
No music	Calm	−1.07	0.287	0.862	−3.06	−0.15	−0.43−0.13
No music	Lively	−2.45	0.015	0.044	−7.15	−0.36	−0.64—0.07
Calm	Lively	−1.40	0.162	0.486	−4.08	−0.20	−0.49−0.08
Self-evaluated learning							
No music	Calm	−2.53	0.012	0.035	−0.38	−0.36	−0.64—0.08
No music	Lively	−1.10	0.270	0.811	−0.17	−0.16	−0.44−0.12
Calm	Lively	1.40	0.163	0.489	0.22	0.20	−0.08−0.49

Note. $df_1 = 2$, $df_2 = 285$; R – retention, T – transfer

The majority of post-hoc comparisons indicate that participants who watched videos with lively background music exhibited slightly higher confidence in their answers, regardless of correctness, compared to those in the group without background music. However, individuals who learned from videos featuring calm music believed they gained slightly more knowledge from the videos compared to the control group. These results again imply that added music positively influences learners' emotions, leading to increased confidence in their learning and subjective test performance, though this effect does not translate into actual test performance.

To summarize, while there were minor differences in subjective learning, no significant differences were found in objective measures of immediate learning, leading us to reject Hypothesis 11. This finding not only contributes to the mixed literature on the effect of background music on learning (de la Mora Velasco and Hirumi, 2020; Kämpfe et al., 2010), but also provides additional context. Previous studies have indicated that music with a faster tempo can have a different effect on learning compared to slower tempo music (Cassidy and MacDonald, 2007; Meyerhoff et al., 2022; Su et al., 2023; Thompson et al., 2011). However, our results indicate no significant differences in the effects of either slow or fast tempo music. One possible explanation is that the learning task was sufficiently challenging, as background music tends to have a stronger effect on easier tasks compared to more difficult ones (Meyerhoff et al., 2022; Su et al., 2023).

Objective and subjective test performance in the delayed part of the experiment

118 (38.44%) participants retook the same knowledge test after seven days. Among them, 40 participants belonged to the control group, 35 had been exposed to videos featuring calm background music, and the remaining 43 had watched videos with lively background music. This section will mirror the previous one, maintaining identical variables, with the exception of the self-evaluated learning variable, which was only part of the main experiment. Descriptive statistics are displayed in Table 111.

Table 111: Descriptive statistics of the learning outcomes from the delayed part of the experiment divided by group

	No music	Calm music	Lively music
Knowledge			
<i>M (SD)</i>	15.51 (4.44)	16.23 (5.20)	16.05 (4.64)
Min–Max	6–26	6–24	6–25
Skewness	0.04	–0.33	0.18
Kurtosis	–0.36	–0.63	–0.61
Retention			
<i>M (SD)</i>	10.05 (3.08)	10.40 (3.77)	10.35 (3.36)
Min–Max	4–16	3–17	4–17
Skewness	–0.29	–0.21	0.26
Kurtosis	–0.40	–0.53	–0.79
Transfer			
<i>M (SD)</i>	5.46 (1.74)	5.83 (1.95)	5.70 (1.70)
Min–Max	2–10	1–9	2–9
Skewness	0.21	–0.25	0.13
Kurtosis	0.13	–0.31	0.01
Certainty in all answers			
<i>M (SD)</i>	57.84 (18.76)	64.78 (21.33)	65.33 (17.60)
Min–Max	5.86–85.69	25.62–99.14	22.76–92.10
Skewness	–0.85	–0.58	–0.67
Kurtosis	0.49	–0.61	0.01
Certainty in correct answers			
<i>M (SD)</i>	60.43 (20.56)	66.92 (22.54)	69.73 (17.69)
Min–Max	3.89–89.17	27.50–100	24.55–96

Skewness	−0.76	−0.71	−0.92
Kurtosis	0.10	−0.79	0.56
Certainty in incorrect answers			
<i>M (SD)</i>	53.45 (16.75)	59.92 (20.96)	59.33 (16.64)
Min–Max	6.75–81.67	24.91–100	20.31–92.57
Skewness	−0.92	−0.09	−0.32
Kurtosis	0.62	−0.68	−0.14
Certainty in all retention answers			
<i>M (SD)</i>	55.81 (18.58)	63.66 (20.93)	63.16 (18.52)
Min–Max	7.89–85.53	25–98.68	20–93.21
Skewness	−0.67	−0.51	−0.56
Kurtosis	−0.02	−0.67	−0.15
Certainty in correct retention answers			
<i>M (SD)</i>	59.28 (20.26)	65.88 (22.79)	67.66 (20.16)
Min–Max	8.75–89.06	25–100	20–97.73
Skewness	−0.62	−0.68	−0.69
Kurtosis	−0.41	−0.88	−0.14
Certainty in incorrect retention answers			
<i>M (SD)</i>	50.38 (16.71)	57.66 (20.54)	56.27 (17.10)
Min–Max	7.67–75	25–100	18.75–95
Skewness	−0.62	0.14	−0.18
Kurtosis	−0.10	−0.58	−0.15
Certainty in all transfer answers			
<i>M (SD)</i>	61.70 (20.95)	66.90 (23.14)	69.46 (17.21)
Min–Max	2–96	25–100	27.50–94.90
Skewness	−0.83	−0.57	−0.80
Kurtosis	0.66	−0.75	0.24
Certainty in correct transfer answers			
<i>M (SD)</i>	63.32 (23.19)	68.14 (23.99)	73.44 (17.70)
Min–Max	0–100	25–100	29.17–100
Skewness	−0.57	−0.66	−0.92
Kurtosis	0.09	−0.81	0.44
Certainty in incorrect transfer answers			
<i>M (SD)</i>	58.06 (20.00)	65.05 (24.42)	65.36 (19.52)
Min–Max	4–96.67	24.67–100	25–100

Skewness	−0.70	−0.32	−0.32
Kurtosis	0.37	−1.05	−0.71
Self-evaluated test performance			
<i>M</i> (<i>SD</i>)	3.33 (1.19)	3.97 (1.36)	3.77 (1.11)
Min–Max	1–6	1–7	1–7
Skewness	−0.19	−0.17	0.38
Kurtosis	−0.65	0.41	1.50

Although the considered variables generally did not satisfy the assumption of normality, neither of them violated the assumption of homogeneity (Appendix 20). Consequently, ANOVAs were conducted to compare the groups, and the results are detailed in Table 112.

Table 112: ANOVA comparisons of the three groups on learning outcomes in the delayed part of the experiment

	<i>F</i>	<i>df</i> ₂	<i>p</i>	η^2p
Knowledge	0.232	114	0.793	0.004
Retention	0.118	114	0.889	0.002
Transfer	0.404	114	0.669	0.007
Certainty	1.897	115	0.155	0.032
Certainty in correct answers	2.271	115	0.108	0.038
Certainty in incorrect answers	1.544	115	0.218	0.026
R Certainty	2.044	115	0.134	0.034
R Certainty in correct answers	1.791	115	0.171	0.030
R Certainty in incorrect answers	1.776	115	0.174	0.030
T Certainty	1.543	115	0.218	0.026
T Certainty in correct answers	2.280	115	0.107	0.038
T Certainty in incorrect answers	1.478	114	0.232	0.025
Self-evaluated test performance	2.831	115	0.063	0.047

Note. *df*₁ = 2; R – retention, T – transfer

There were no significant differences in any of the learning outcomes seven days after watching the videos, except for the self-evaluation of participants' test performance. Post-hoc comparisons for this variable are represented in Table 113.

Table 113: Post-hoc comparisons for Self-evaluated test performance

Experimental groups		<i>t</i>	<i>p</i>	<i>p</i> _{Bonferroni}	Mean difference	<i>d</i>	95% <i>CI</i>
No music	Calm	−2.30	0.023	0.070	−0.65	−0.53	−1.00—0.07
No music	Lively	−1.66	0.100	0.299	−0.44	−0.36	−0.80—0.07
Calm	Lively	0.74	0.462	1.000	0.20	0.17	−0.28—0.62

Note. *df* = 115

A marginally significant distinction surfaced between the control and calm music groups, as the latter perceived that they performed slightly better on the test compared to their counterparts who viewed videos without additional elements. However, this perception was not reflected in the objective test results.

Similarly to the results section on the immediate part of the experiment, a MANCOVA was performed also with some learning outcomes from the delayed part, controlling for the same covariates, but this time including five dependent variables: delayed retention, transfer, certainty in retention answers, certainty in transfer answers, and self-evaluated test performance. No significant effect was detected (Wilks' Lambda = 0.94, $F(10, 198) = 0.64$, $p = 0.774$; $\chi^2(30) = 31.58$, $p = 0.387$, $W = 0.97$, $p = 0.016$).

Table 114: ANCOVA comparisons with eleven covariates on learning variables in the delayed part of the experiment

	ANCOVA*			Homogeneity test**		Normality test	
	<i>F</i>	<i>p</i>	η^2p	<i>F</i>	<i>p</i>	<i>W</i>	<i>p</i>
Knowledge	0.91	0.406	0.02	1.79	0.172	0.98	0.109
Retention	0.71	0.493	0.01	0.59	0.553	0.98	0.217
Transfer	0.90	0.411	0.02	1.47	0.235	0.99	0.697
Certainty	2.84	0.063	0.05	0.21	0.808	0.96	0.001
Certainty in correct answers	2.91	0.059	0.05	0.18	0.838	0.95	< .001
Certainty in incorrect answers	2.48	0.089	0.05	0.47	0.626	0.99	0.232
R Certainty	3.07	0.051	0.06	0.31	0.735	0.96	0.002
R Certainty in correct answers	2.49	0.088	0.05	0.41	0.663	0.95	< .001

R Certainty in incorrect answers	2.67	0.074	0.05	0.68	0.508	0.99	0.680
T Certainty	2.17	0.119	0.04	0.68	0.507	0.98	0.030
T Certainty in correct answers	2.52	0.086	0.05	0.36	0.697	0.98	0.056
T Certainty in incorrect answers	2.02	0.138	0.04	0.92	0.403	0.98	0.105
Self-evaluated test performance	3.35	0.039	0.06	0.17	0.844	0.99	0.849

Note. * $df_1 = 2$, $df = 104$; ** $df_1 = 2$, $df_2 = 115$; R – retention, T – transfer

Results of ANCOVAs for all delayed learning outcomes are displayed in Table 114. Similarly to the results from the immediate part of the experiment, there were some marginally significant differences in the subjective outcomes, which were then compared using post-hoc tests presented in Table 115.

Table 115: ANCOVA post-hoc comparisons for some of the learning outcome variables in the delayed part of the experiment

Experimental groups		<i>t</i>	<i>p</i>	<i>p</i> _{Bonferroni}	Mean difference	<i>d</i>	95% CI
Certainty							
No music	Calm	-2.05	0.043	0.130	-8.68	-0.49	-0.97—0.01
No music	Lively	-2.08	0.040	0.119	-8.89	-0.50	-0.99—0.02
Calm	Lively	-0.05	0.961	1.000	-0.21	-0.01	-0.49—0.46
Certainty in correct answers							
No music	Calm	-1.87	0.064	0.191	-8.42	-0.45	-0.93—0.03
No music	Lively	-2.25	0.026	0.079	-10.19	-0.54	-1.03—0.06
Calm	Lively	-0.39	0.695	1.000	-1.77	-0.09	-0.57—0.38
Certainty in incorrect answers							
No music	Calm	-1.94	0.055	0.165	-8.05	-0.47	-0.95—0.01
No music	Lively	-1.92	0.058	0.174	-8.00	-0.46	-0.95—0.02
Calm	Lively	0.01	0.990	1.000	0.05	0.00	-0.47—0.48
R Certainty							
No music	Calm	-2.25	0.027	0.081	-9.48	-0.54	-1.02—
No music	Lively	-2.04	0.044	0.132	-8.65	-0.49	-0.98—
Calm	Lively	0.20	0.845	1.000	0.83	0.05	-0.43—

R Certainty in correct answers							
No music	Calm	-1.82	0.071	0.213	-8.46	-0.44	-0.92—0.04
No music	Lively	-2.03	0.045	0.136	-9.45	-0.49	-0.97—0.01
Calm	Lively	-0.21	0.831	1.000	-0.99	-0.05	-0.53—0.42
R Certainty in incorrect answers							
No music	Calm	-2.11	0.038	0.113	-8.70	-0.51	-0.99—0.02
No music	Lively	-1.88	0.062	0.187	-7.82	-0.46	-0.94—0.03
Calm	Lively	0.21	0.833	1.000	0.87	0.05	-0.43—0.53
T Certainty in correct answers							
No music	Calm	-1.40	0.163	0.490	-6.99	-0.34	-0.82—0.14
No music	Lively	-2.22	0.029	0.086	-11.12	-0.54	-1.02—0.05
Calm	Lively	-0.83	0.409	1.000	-4.13	-0.20	-0.68—0.28
Self-evaluated test performance*							
No music	Calm	-2.54	0.013	0.038	-0.69	-0.61	-1.09—0.13
No music	Lively	-1.72	0.089	0.266	-0.47	-0.41	-0.90—0.07
Calm	Lively	0.81	0.420	1.000	0.22	0.19	-0.28—0.67

Note. $df = 144$; * $df = 103$; R – retention, T – transfer

The outcomes mirrored the pattern observed in the main phase of the experiment. Both the calm music and the lively music groups displayed slightly elevated levels of certainty in their answers, though these differences did not withstand the Bonferroni correction. However, there was one significant difference in one aspect – the calm music group exhibited significantly higher ratings of self-evaluated test performance compared to the group without added music. Therefore, we can conclude that the positive effect of adding background music on subjective learning and self-evaluated test performance appears to be not only immediate but also long-term, which is a novel finding since no other studies reported on the role of music on subjective learning. However, this effect does not translate into actual learning, so our main hypothesis (H11) was disproven.

Comparison of objective and subjective test performance between sessions

To examine the evolution of learning variables over time, pairwise Student's t -tests were used to compare the same variables in both immediate and delayed conditions. High correlations between the test results and self-evaluated test performance in both testing sessions were observed ($r_{knowledge} = 0.850$, $p < .001$; $r_{retention} = 0.788$, $p < .001$; $r_{transfer} = 0.749$, $p < .001$; $r_{self-evaluation} = 0.772$, $p < .001$). Table 116 displays descriptive statistics of results from each testing session.

Table 116: Descriptive statistics of the learning outcomes from the immediate ($N = 304$) and delayed part of the experiment ($N = 118$)

	<i>M</i>		<i>SD</i>		Min–Max		Skewness		Kurtosis	
	Imm	Del	Imm	Del	Imm	Del	Imm	Del	Imm	Del
K	17.08	15.92	4.75	4.72	0–28	6–26	–0.09	–0.03	–0.17	–0.60
R	11.35	10.26	3.53	3.37	0–18	3–17	–0.00	–0.04	–0.36	–0.57
T	5.73	5.66	1.75	1.78	0–10	1–10	–0.29	0.03	0.05	–0.20
C	67.27	62.63	16.28	19.30	0–99.14	5.86–99.14	–0.80	–0.64	0.48	–0.08
C ^y	70.33	65.74	16.62	20.45	0–100	3.89–100	–0.92	–0.78	0.52	–0.13
C ⁿ	61.58	57.51	16.29	18.14	0–100	6.75–100	–0.60	–0.30	0.35	–0.02
RC	66.85	60.82	16.73	19.46	0–98.68	7.89–98.68	–0.73	–0.51	0.29	–0.33
RC ^y	70.03	64.29	17.46	21.15	0–100	8.75–100	–0.86	–0.61	0.35	–0.56
RC ⁿ	59.65	54.69	17.17	18.18	0–100	7.67–100	–0.44	–0.09	0.10	–0.13
TC	68.08	66.07	17.76	20.47	0–100	2–100	–0.85	–0.74	0.38	0.09
TC ^y	70.79	68.44	18.79	21.84	0–100	0–100	–0.87	–0.75	0.22	–0.09
TC ⁿ	64.07	62.84	19.34	21.34	0–100	4–100	–0.48	–0.37	–0.09	–0.47
SE	3.79	3.68	1.13	1.23	1–7	1–7	–0.03	0.03	0.30	0.34

Note. Imm – immediate part of the experiment, Del – delayed part of the experiment; K – knowledge, R – retention, T – transfer, C – certainty, ^y – correct answers, ⁿ – incorrect answers, SE – self-evaluation

While any of the variables did not meet the assumption of normality, Student's *t*-test was still performed due to the sufficiently large sample size. As can be deducted from Tables 116 and 117, there was a significant decrease in total test performance and the retention rate, as well as general certainty rate and certainty level in retention-related questions in seven days. The difference in self-evaluated test performance, transfer, and level of certainty in transfer questions, however, remained constant.

Table 117: Pairwise comparisons of the learning variables in the immediate and delayed parts of the experiment

Variable	<i>t</i>	<i>p</i>	<i>Mean difference</i>	<i>95% CI</i>	<i>d</i>	<i>95% CI</i>
Knowledge*	4.82	< .001	1.15	0.68–1.63	0.45	0.25–0.63
Retention*	5.21	< .001	1.09	0.67–1.50	0.48	0.29–0.67
Transfer*	0.59	0.555	0.07	–0.16–0.30	0.05	–0.13–0.24
Certainty [†]	4.59	< .001	4.64	2.64–6.65	0.42	0.23–0.61
Certainty in correct answers [†]	3.85	< .001	4.58	2.23–6.94	0.35	0.17–0.54
Certainty in incorrect answers [†]	3.50	< .001	4.07	1.76–6.37	0.32	0.14–0.51
R Certainty [†]	5.39	< .001	6.03	3.82–8.25	0.50	0.30–0.69
R Certainty in correct answers [†]	4.46	< .001	5.74	3.19–8.28	0.41	0.22–0.60
R Certainty in incorrect answers [†]	3.79	< .001	4.96	2.37–7.55	0.35	0.16–0.53
T Certainty [†]	1.78	0.078	2.01	–0.23–4.24	0.16	–0.02–0.34
T Certainty in correct answers [†]	1.67	0.098	2.36	–0.44–5.16	0.15	–0.03–0.33
T Certainty in incorrect answers*	0.88	0.381	1.24	–1.55–4.02	0.08	–0.10–0.26
Self-evaluation [†]	1.49	0.139	0.11	–0.04–0.26	0.14	–0.04–0.32

Note. **df* = 116, [†]*df* = 117; R – retention, T – transfer

3.5.3.6 Additional analyses

Comparisons based on English proficiency

Following the example of Study 1, we examined potential differences among groups based on varying levels of English proficiency to ascertain whether the results exhibit variations across different English competency levels. Given that only approximately 7% of the sample constituted native English speakers, we once again used the LexTALE score to categorize participants based on their language proficiency. For this study, we set the LexTALE test score threshold at 69 (higher than the threshold of 63 used in Study 1), with 49.50% (148 participants) scoring below this threshold and 50.50% (151

participants) scoring above it. Similar to Study 1, the results will be presented separately for the low and high-proficiency groups.

Lower proficiency group

Appendices 21 and 22 display the descriptives and ANCOVA results for participants scoring below 69 on LexTALE. Here, we will present post-hoc comparisons of variables that showed a (marginally) significant main effect. These can be viewed in Table 118.

Table 118: ANCOVA post-hoc comparisons for the lower proficiency group – Study 2

Experimental groups		<i>t</i>	<i>p</i>	<i>p</i> _{Bonferroni}	Mean difference	<i>d</i>	95% <i>CI</i>
Perceived video pleasantness (<i>df</i> = 134)							
No music	Calm	-2.96	0.004	0.011	-0.72	-0.61	-1.02—0.20
No music	Lively	-0.85	0.394	1.000	-0.22	-0.18	-0.61—0.24
Calm	Lively	1.98	0.049	0.148	0.51	0.42	-0.00—0.85
Negative activation (<i>df</i> = 134)							
No music	Calm	2.06	0.042	0.125	0.39	0.42	0.01—0.83
No music	Lively	-0.83	0.406	1.000	-0.17	-0.18	-0.60—0.25
Calm	Lively	-2.81	0.006	0.017	-0.56	-0.60	-1.03—0.71
Valence (PANAVA-KS; <i>df</i> = 134)							
No music	Calm	-2.33	0.022	0.065	-0.46	-0.48	-0.89—0.07
No music	Lively	-0.83	0.408	1.000	-0.17	-0.18	-0.60—0.25
Calm	Lively	1.40	0.164	0.493	0.29	0.30	-0.13—0.72
Valence (<i>df</i> = 145)*							
No music	Calm	-3.19	0.002				
No music	Lively	-1.68	0.096				
Calm	Lively	1.46	0.147				
Paying attention (<i>df</i> = 134)							
No music	Calm	-2.40	0.018	0.053	-0.60	-0.49	-0.90—0.08
No music	Lively	-0.50	0.620	1.000	-0.13	-0.11	-0.53—0.32
Calm	Lively	1.81	0.073	0.220	0.47	0.39	-0.04—0.81
Exerting more effort (<i>df</i> = 134)							
No music	Calm	-0.52	0.602	1.000	-0.16	-0.11	-0.51—0.30
No music	Lively	1.80	0.074	0.221	0.57	0.39	-0.04—0.81
Calm	Lively	2.31	0.023	0.068	0.72	0.49	0.07—0.92

Enjoyment ($df = 134$)							
No music	Calm	-2.34	0.021	0.063	-0.58	-0.48	-0.89—0.07
No music	Lively	-0.85	0.395	1.000	-0.22	-0.18	-0.61—0.24
Calm	Lively	1.39	0.168	0.505	0.36	0.30	-0.13—0.72
More lessons like this ($df = 134$)							
No music	Calm	-3.20	0.002	0.005	-0.88	-0.66	-1.07—0.24
No music	Lively	-1.40	0.165	0.496	-0.40	-0.30	-0.72—0.13
Calm	Lively	1.67	0.098	0.294	0.48	0.36	-0.07—0.78
Mental effort (average; $df = 134$)							
No music	Calm	-1.60	0.112	0.336	-0.50	-0.33	-0.74—0.08
No music	Lively	1.20	0.231	0.694	0.39	0.26	-0.17—0.68
Calm	Lively	2.74	0.007	0.021	0.89	0.59	0.16—1.01
Certainty in correct retention answers ($df = 134$)							
No music	Calm	-0.93	0.354	1.000	-3.73	-0.19	-0.60—0.22
No music	Lively	-2.28	0.024	0.073	-9.54	-0.49	-0.92—0.06
Calm	Lively	-1.39	0.167	0.501	-5.81	-0.30	-0.72—0.13
Self-evaluated learning ($df = 134$)							
No music	Calm	-2.99	0.003	0.010	-0.61	-0.61	-1.02—0.20
No music	Lively	-1.68	0.096	0.288	-0.36	-0.36	-0.79—0.07
Calm	Lively	1.18	0.239	0.716	0.25	0.25	-0.17—0.68

Note. *Quade's test results reported

From Table 118 we can see that some differences between groups are more pronounced in the lower proficiency group compared to the whole sample. Namely, there were (marginally) significant differences between the calm music and no music groups in perceived video pleasantness, valence (both measures), paying attention to the videos, enjoyment, wish to have more similar lessons, and self-evaluated learning. Specifically, participants with lower language proficiency who watched videos with added calm background music perceived the videos as more pleasant and enjoyed them more, paid more attention to the content, felt they learned more, and expressed a higher desire for similar lessons in the future compared to the control group, highlighting the positive effect of calm music on learners' affective states. Between the lively and no music group, there were marginal differences in valence and certainty in correct retention answers, while the calmer music group had significantly different scores from the lively music group in negative activation, exerting more effort, and mental effort. To be exact, learners with lower proficiency in the video language who watched videos with calm music

reported a greater decrease in negative activating emotions, put more effort into learning, and experienced higher mental effort compared to those in the lively music condition.

These results show that when language proficiency is lower, calm background music demands more cognitive resources than lively music, but it provides more enjoyment and motivates the learners to pay more attention to the content compared to videos without background music. In other words, for individuals who lack strong proficiency in the foreign language, calm background music may benefit their emotional state but can negatively impact their mental effort and attention, especially when compared to more energetic music. This finding is inconsistent with previous research, which suggests that calm, slow-tempo music allows for better recovery from acoustic interference and therefore demands fewer cognitive resources compared to lively music (Cassidy and MacDonald, 2007; Thompson et al., 2011). In contrast, lively, higher-tempo music has been shown to improve performance more than calm music (Su et al., 2023), suggesting that these effects may depend on learners' individual differences.

Higher proficiency group

On the other hand, the results of the higher English proficiency group can be seen in Appendices 21 and 23, while the post-hoc comparisons of variables with significant main effects are presented in Table 119.

Table 119: ANCOVA post-hoc comparisons for the higher proficiency group – Study 2

Experimental groups		<i>t</i>	<i>p</i>	<i>p</i> _{Bonferroni}	Mean difference	<i>d</i>	95% <i>CI</i>
Interest (delayed, <i>df</i> = 137)							
No music	Calm	-1.61	0.113	0.340	-0.69	-0.61	-1.37–0.16
No music	Lively	-2.19	0.033	0.100	-0.87	-0.76	-1.48–0.05
Calm	Lively	-0.42	0.675	1.000	-0.18	-0.16	-0.90–0.59
Exerting more effort (<i>df</i> = 137)							
No music	Calm	-0.48	0.634	1.000	-0.13	-0.10	-0.51–0.31
No music	Lively	-2.68	0.008	0.025	-0.71	-0.55	-0.97–0.14
Calm	Lively	-2.22	0.028	0.085	-0.59	-0.45	-0.86–0.05
Mental effort (average) (<i>df</i> = 148)*							
No music	Calm	-1.67	0.098				
No music	Lively	-2.42	0.017				
Calm	Lively	-0.74	0.461				

Self-evaluated test performance ($df = 137$)							
No music	Calm	-2.27	0.025	0.075	-0.51	-0.47	-0.88—0.06
No music	Lively	-1.56	0.120	0.361	-0.35	-0.32	-0.73—0.09
Calm	Lively	0.70	0.483	1.000	0.16	0.14	-0.26—0.55
Certainty in correct answers (delayed, $df = 47$)							
No music	Calm	-1.97	0.055	0.164	-13.33	-0.74	-1.51—0.03
No music	Lively	-2.00	0.052	0.155	-12.48	-0.69	-1.41—0.02
Calm	Lively	0.13	0.898	1.000	0.85	0.05	-0.69—0.79
Certainty in retention answers (delayed, $df = 47$)							
No music	Calm	-2.02	0.049	0.148	-13.43	-0.76	-1.53—0.01
No music	Lively	-1.81	0.077	0.231	-11.12	-0.63	-1.34—0.08
Calm	Lively	0.35	0.724	1.000	2.31	0.13	-0.61—0.87
Certainty in correct retention answers (delayed, $df = 47$)							
No music	Calm	-2.03	0.048	0.144	-14.36	-0.76	-1.53—0.01
No music	Lively	-1.94	0.058	0.174	-12.69	-0.67	-1.39—0.04
Calm	Lively	0.24	0.811	1.000	1.67	0.09	-0.65—0.83
Certainty in incorrect transfer answers (delayed, $df = 47$)							
No music	Calm	-1.98	0.053	0.159	-15.40	-0.77	-1.56—0.03
No music	Lively	-1.94	0.058	0.175	-14.10	-0.70	-1.45—0.04
Calm	Lively	0.18	0.861	1.000	1.30	0.06	-0.68—0.81
Self-evaluated test performance (delayed; $df = 58$)*							
No music	Calm	-2.34	0.023				
No music	Lively	-1.88	0.066				
Calm	Lively	0.59	0.560				

Note. *Quade's test results reported

The variables outlined in Table 119 differ significantly from those demonstrating a main effect in the lower proficiency group. In the higher proficiency group, participants exposed to lively music exhibited significantly greater effort in learning the material compared to those with no music, while the remaining variables displayed only marginal differences. Noteworthy distinctions included variations between the no music and lively music conditions in interest during the second part of the experiment, mental effort, and delayed self-evaluated test performance. Participants who watched videos with lively music reported higher delayed interest, better delayed subjective test performance, greater mental effort, and more exerted effort compared to the control group. Additionally, the calm music group reported better subjective test performance after both

the immediate and delayed tests and exerted more effort compared to the no music group. Finally, there was a difference between the calm and lively music groups in their exertion of effort during learning – the lively music group reported higher effort exertion than the calm music group.

Unlike the lower proficiency group, livelier music in this instance appeared to demand more cognitive resources than no or calm music. This aligns with previous research findings suggesting that high-tempo music can be more detrimental to learning than slow-tempo music, as it includes a higher number of auditory events per unit of time, consuming more of the listeners' attention (Cassidy and MacDonald, 2007; Thompson et al., 2011). As mentioned before, the effect of type (tempo) of background music may depend on learners' characteristics, such as language proficiency.

Comparisons based on study program

In contrast to Study 1, separate ANCOVAs were also performed based on the participants' study program. We were interested in whether there would be any differences between the groups if we looked separately at students from fields of education related to wood science, engineering, and forestry (KLASIUS-P-1–16 groups: Engineering, manufacturing and construction and Agriculture, forestry, fisheries and veterinary, $n = 106$), in short, study programs that are more closely aligned to the learning content of the videos, and participants who studied in a completely non-relevant field ($n = 194$). As seen in the subchapter on group differences, these two groups differed significantly in prior knowledge, experience, interest, and number of spoken languages. Given these distinctions, the impact of background music on these participant groups may differ. Despite the significant difference in size, with the first group comprising 35.33% (106) of participants and the second group 64.57% (194), the findings may yield some insights.

Study programs more closely aligned wood science

Descriptive statistics and ANCOVA/Quade's test results are displayed in Appendices 24 and 25. From Table 138 in Appendix 25 it can be clearly seen that the only main effects found in this cohort are the ones related to the level of confidence in one's answers. Table 120 reports post-hoc comparisons of these variables.

Table 120: ANCOVA post-hoc comparisons for the wood science related educational programs

Experimental groups		<i>t</i>	<i>p</i>	<i>p</i> _{Bonferroni}	Mean difference	<i>d</i>	95% <i>CI</i>
Certainty							
No music	Calm	-1.15	0.254	0.761	-4.22	-0.29	-0.79—0.21
No music	Lively	-3.07	0.003	0.008	-11.21	-0.77	-1.28—0.26
Calm	Lively	-1.91	0.059	0.176	-6.99	-0.48	-0.98—0.02
Certainty in correct answers							
No music	Calm	-1.52	0.133	0.399	-5.79	-0.38	-0.89—0.12
No music	Lively	-2.93	0.004	0.013	-11.13	-0.73	-1.24—0.22
Calm	Lively	-1.40	0.164	0.491	-5.34	-0.35	-0.85—0.15
Certainty in incorrect answers							
No music	Calm	-1.27	0.208	0.623	-4.78	-0.32	-0.82—0.18
No music	Lively	-3.33	0.001	0.004	-12.49	-0.83	-1.35—0.32
Calm	Lively	-2.05	0.043	0.129	-7.70	-0.51	-1.02—0.01
Certainty in retention answers							
No music	Calm	-0.80	0.423	1.000	-3.07	-0.20	-0.70—0.30
No music	Lively	-2.76	0.007	0.021	-10.49	-0.69	-1.20—0.18
Calm	Lively	-1.95	0.054	0.162	-7.42	-0.49	-0.99—0.01
Certainty in correct retention answers							
No music	Calm	-1.38	0.172	0.515	-5.61	-0.35	-0.85—0.16
No music	Lively	-2.66	0.009	0.028	-10.77	-0.67	-1.17—0.16
Calm	Lively	-1.27	0.206	0.619	-5.16	-0.32	-0.82—0.18
Certainty in incorrect retention answers							
No music	Calm	-0.77	0.445	1.000	-3.16	-0.19	-0.69—0.31
No music	Lively	-2.92	0.004	0.013	-11.92	-0.73	-1.24—0.22
Calm	Lively	-2.14	0.035	0.105	-8.77	-0.54	-1.04—0.03
Certainty in transfer answers*							
No music	Calm	-1.01	0.317				
No music	Lively	-2.58	0.011				
Calm	Lively	-1.60	0.112				
Certainty in correct transfer answers							
No music	Calm	-1.49	0.139	0.416	-6.40	-0.38	-0.88—0.13
No music	Lively	-2.95	0.004	0.012	-12.58	-0.74	-1.25—0.23

Calm	Lively	−1.45	0.151	0.452	−6.18	−0.36	−0.86–0.14
Certainty in incorrect transfer answers							
No music	Calm	−1.26	0.212	0.637	−5.87	−0.32	−0.82–0.19
No music	Lively	−2.55	0.013	0.038	−11.83	−0.64	−1.14–0.13
Calm	Lively	−1.28	0.203	0.610	−5.96	−0.32	−0.82–0.18

Note. $df = 92$, *Quade's test results reported ($df = 103$)

Results show that when looking only at the students from educational fields relevant to wood science, the only difference between groups was in the level of certainty in their answers on the whole knowledge test. Specifically, the group that watched videos with lively background music exhibited a higher level of certainty in their answers compared to the group with no background music added. This holds for all certainty-related variables in the immediate knowledge test but was not repeated a week after the initial experiment.

Study programs not related to wood science

When focusing only on students that who did not study a field that is at least somehow relevant to the learning videos content, we can see that there are many more differences present (Appendices 24 and 26). In this section, Table 121 represents post-hoc tests that were made for variables that exhibited significant main effects when performing ANCOVA.

Table 121: ANCOVA post-hoc comparisons for the educational programs not related to wood science related

Experimental groups		t	p	$p_{Bonferroni}$	Mean difference	d	95% CI
Pleasantness							
No music	Calm	−2.94	0.004	0.011	−0.65	−0.52	−0.88–0.17
No music	Lively	0.47	0.638	1.000	0.10	0.08	−0.27–0.44
Calm	Lively	3.32	0.001	0.003	0.75	0.61	0.24–0.98
Valence (PANAVA–KS)							
No music	Calm	−2.28	0.024	0.072	−0.38	−0.41	−0.76–0.05
No music	Lively	−1.36	0.176	0.529	−0.23	−0.24	−0.60–0.11
Calm	Lively	0.89	0.376	1.000	0.15	0.16	−0.20–0.52
Valence							

No music	Calm	-3.08	0.002	0.007	-0.59	-0.55	-0.91—0—19
No music	Lively	-1.36	0.175	0.524	-0.26	-0.24	-0.60—0—11
Calm	Lively	1.68	0.095	0.284	0.33	0.31	-0.06—0—67
Situational interest							
No music	Calm	-2.49	0.014	0.041	-0.41	-0.44	-0.80—0—09
No music	Lively	0.42	0.678	1.000	0.07	0.07	-0.28—0—43
Calm	Lively	2.83	0.005	0.016	0.48	0.52	0.15—0—88
Intrinsic motivation							
No music	Calm	-2.51	0.013	0.039	-0.44	-0.45	-0.80—0—09
No music	Lively	-0.40	0.690	1.000	-0.07	-0.07	-0.43—0—28
Calm	Lively	2.05	0.042	0.126	0.37	0.38	0.01—0—74
Paying attention							
No music	Calm	-3.13	0.002	0.006	-0.72	-0.56	-0.92—0—20
No music	Lively	-0.71	0.478	1.000	-0.16	-0.13	-0.48—0—23
Calm	Lively	2.35	0.020	0.059	0.55	0.43	0.07—0—80
More lessons like this							
No music	Calm	-3.05	0.003	0.008	-0.75	-0.54	-0.90—0—19
No music	Lively	-1.13	0.259	0.776	-0.28	-0.20	-0.56—0—15
Calm	Lively	1.86	0.065	0.195	0.47	0.34	-0.02—0—70
Self-evaluated learning							
No music	Calm	-2.76	0.006	0.019	-0.53	-0.49	-0.85—0—14
No music	Lively	-1.06	0.291	0.872	-0.20	-0.19	-0.54—0—16
Calm	Lively	1.65	0.101	0.304	0.32	0.30	-0.06—0—67
Self-evaluated test performance (delayed)*							
No music	Calm	-2.29	0.025	0.075	-0.77	-0.66	-1.25—0—08
No music	Lively	-2.22	0.030	0.089	-0.74	-0.64	-1.22—0—06
Calm	Lively	0.10	0.924	1.000	0.03	0.03	-0.54—0—59

Note. $df = 179$, * $df = 73$

In the cohort of students who were from educational fields not at all related to wood science, participants from the group with calm background music perceived the videos as significantly more pleasant, had higher situational interest, and were more motivated to pay attention to the videos compared to participants from the other two groups. In addition, compared to the group without any music, they also expressed higher valence, intrinsic motivation, wished for more similar lessons, and perceived that they learned

more from the videos. These results show that the addition of calm background music had a beneficial effect on learners' emotional state, and that this effect is more pronounced for students who were less familiar with the video topic.

The role of personality characteristics

In this study, a comprehensive analysis was conducted through ANCOVAs, incorporating 11 covariates to examine their impact on the main outcome variables. Of particular interest were the big five personality characteristics, which comprised five of the 11 covariates. To present the influence of these personality traits on the primary outcomes, Table 122 provides a focused summary exclusively showcasing the ANCOVA results for the big five personality characteristics.

Table 122: Big five personality characteristics as covariates on main outcome variables

	Openness	Conscientiousness	Extroversion	Agreeableness	Neuroticism
VVA	0.10	0.24	2.19	3.74*	0.00
VAC	0.39	1.92	0.09	0.91	0.97
PA	0.45	4.53*	0.75	4.30*	0.01
NA	0.72	2.48	0.00	3.88*	3.01 [†]
VA	2.28	1.39	0.27	19.56***	1.68
ACT	0.24	2.99 [†]	1.65	6.18*	0.02
VAL	0.37	0.07	0.13	8.10**	0.12
INT	1.12	1.79	1.97	1.96	0.70
IND	2.19	0.59	0.53	0.14	0.34
MOT	4.96	0.02*	6.05	2.72*	0.27
ATT	2.02	2.58	2.17	9.21**	1.29
DIF	9.03**	0.53	0.95	3.97*	0.25
EFF	1.36	0.01	4.62*	10.46***	5.71*
ENJ	7.85	0.02	5.34*	5.97*	1.49
LES	0.61	1.82	7.99**	2.34	0.00
ICL	3.45	0.68	0.00	1.17	2.45
ECL	3.11 [†]	1.51	0.09	1.04	1.16
GCL	0.44	5.56*	0.61	9.57**	2.24
ME	0.05	0.63	0.42	4.95*	4.46*
SEL	6.10*	1.80	0.68	0.08	2.09

K	2.71	0.15	4.73*	2.43	0.48
R	5.76*	0.02	6.21	1.25	0.63
T	0.12	0.44	0.53	2.85 [†]	0.06
SE	0.40	2.60	0.75	0.39	0.82
KD	0.33	0.46	5.64*	0.04	0.98
RD	0.31	0.31	4.92*	0.44	0.88
TD	0.18	0.44	3.40 [†]	2.93 [†]	0.55
SED	0.79	1.24	0.83	0.01	2.14

Note. [†] $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; VVA – video valence, VAC – video activation level, PA – positive activation change score, NA – negative activation change score, VA – valence (PANAVA-KS) change score, ACT – activation level average change score, VAL – valence average change score, INT – situational interest, IND – interest (delayed), MOT – intrinsic motivation, ATT – paying attention, DIF – difficulty, EFF – effort, ENJ – enjoyment, LES – more lessons, ICL – intrinsic cognitive load, ECL – extraneous cognitive load, GCL – germane cognitive load, ME – general mental effort, SEL – self-evaluated learning, K – knowledge, R – retention, T – transfer, SE – self-evaluation, KD – knowledge (delayed), RD – retention (delayed), TD – transfer (delayed), SED – self-evaluation (delayed)

As can be seen from the table above, agreeableness was the personality characteristic that most often contributed to the main outcomes, particularly in emotional and cognitive variables. Notably, agreeableness significantly contributed to participants' emotional state, motivation, perception of lesson difficulty, attention, enjoyment, germane cognitive load, and (mental) effort. Extroversion also played a role in perceived effort, enjoyment, desire for more lessons of this nature, and overall learning across both segments of the experiment. Conversely, openness, conscientiousness, and neuroticism exhibited a lesser and more sporadic impact on the main outcomes of the experiment.

3.5.3.6 Limitations and implications

The goals of Study 2 were to explore the impact of embedded background music on learners' cognitive, affective, and learning outcomes. Unlike previous research, this study included a control group who watched videos without music and two experimental groups with different types of music – calm and lively. We also examined the influence of English proficiency and the relevance of the study program to the video's content. Based on our results, the addition of any kind of background music does not have an effect on cognitive

and learning outcomes. Adding calm background music, however, can benefit participants' emotional state by helping regulate negative activating emotions, reducing stress or anxiety, and thus creating a more relaxed state conducive to learning. Those who viewed videos with calming music also perceived the lessons as more pleasant, enjoyable, interesting, and were more confident in their test performance. These results were particularly evident among participants with lower English proficiency and those from study programs unrelated to the video content. When examining the sample based on language proficiency, a negative effect of calm music on mental effort was observed in those with lower proficiency, while a negative effect of lively music on mental effort was seen in those with higher proficiency. This suggests that multimedia learning principles may not uniformly apply to all learners.

Like in Study 1, the null results prevent us from drawing definitive conclusions. Despite a sample of over 300 participants with varying levels of content familiarity and English proficiency, each experimental group only had around 100 participants. This size may be adequate for detecting medium to large effects, but not smaller ones, limiting the study's power and generalizability, especially in the delayed part of the experiment where there were even less participants. Background music is a minor addition to learning videos, so its effects on student outcomes may also be minor. While in the pre-studies music affected participants' activation levels, in the main experiment, where music was in the background and participants did not focus on the songs, its impact diminished. While it changed negative activation emotions, it did not affect activation levels as in the pre-study, indicating different effects when music is the focus versus in the background. A larger sample would thus help detect smaller differences. While we included students from programs likely to be interested in the educational material and individuals with varying levels of English proficiency, including native speakers, the sample remained unbalanced. Only a small portion of the sample consisted of native English speakers. Samples in future studies should be larger, more balanced and representative.

Future studies should also use a variety of songs to generalize findings. Although we selected songs validated in pre-studies, the chosen music might not have been equally appealing to all participants, potentially influencing their responses. Personal music preferences were not considered. Our study provides insights into participants' music listening habits while studying and their perception of the chosen songs. The lively song, in particular, was perceived as repetitive, which could have been more distracting than in realistic learning scenarios. Previous research has shown that songs chosen by participants themselves yield better outcomes compared to those provided by

researchers (Lynar et al., 2017). In the future, it would be interesting to compare the effects of music embedded in videos with songs chosen by participants themselves.

Additionally, while we tested participants' knowledge two times – immediately after learning and a week after, the study's short-term nature might not capture long-term effects, so later testing would be beneficial. Furthermore, participants might not have been as motivated as they would be in a real academic setting, as the study material was not part of their coursework and had no real consequences.

Overall, our study adds to the literature on background music in learning videos, with methodological improvements like longer videos, realistic music types, and differentiation between calm and lively music. The controlled experimental setting may not reflect real-world learning environments where various external factors affect the effectiveness of background music on learning outcomes.

The results of adding background music to educational videos are not straightforward, so future studies should build on these results and provide more insight into how different types of songs affect various outcomes and how they interact with various individual differences. They could also use psychophysiological tools for more objective insights into participants' emotional and cognitive processing.

4 CONCLUSIONS

The construction industry is a major economic sector that faces significant environmental challenges, with buildings accounting for about 40% of energy consumption and nearly half of CO₂ emissions in the EU (Bonoli et al., 2021; Clarke and Sahin-Dikmen, 2020). Sustainable practices are therefore essential to mitigate these impacts. Wood is a renewable material with a lower carbon footprint, excellent functional properties, versatility, and has positive effects on people's health and well-being, making it a promising option for sustainable building. However, widespread adoption is hindered by misconceptions and a lack of knowledge. Therefore, educational initiatives for professionals and the public are crucial to creating a favourable market for wood and encouraging its broader use in construction, thereby promoting greater sustainability.

For this reason, we created a series of educational videos on wood as a building material. Learning videos are a particularly impactful online learning tool that enhance comprehension and retention by offering visual and auditory experiences, making educational content more engaging and accessible, especially in challenging or unfamiliar fields (Peters and Romero, 2019; Steffens, 2015). However, not all learning videos and multimedia learning materials are equally effective. Despite an increasing body of evidence identifying effective educational video design principles, much remains unknown. Research grounded in cognitive load theory (Sweller, 1994) and the cognitive theory of multimedia learning (Mayer, 2014) seeks to optimize these materials. Recent advancements include incorporating emotional design principles to improve learning by influencing learners' emotions (Plass and Kaplan, 2016). This has led to more comprehensive theories like the cognitive-affective theory of learning with media (Moreno, 2006; Moreno and Mayer, 2007), and the cognitive affective model of e-learning (Lawson et al., 2021b; Mayer, 2020), which formed a theoretical basis for our studies.

The goal of the dissertation was to investigate how a minimal manipulation, such as the narrator's emotional tone, and a seductive detail such as background music, affect learners' affective, cognitive, and learning outcomes. Given that most research on multimedia learning has been conducted in native languages, we targeted non-native English speakers to assess how these emotional design elements perform in a foreign language setting. Additionally, we evaluated the effectiveness of same-language subtitles for enhancing learning from videos presented in a non-native language.

For this purpose, we conducted three pre-studies to validate the used materials and two main experiments. Study 1 involved a 2x2 between-subjects design with 229 university students who watched educational videos and responded to a series of questions and questionnaires. The experimental factors were the narrator's emotional state (enthusiastic vs. calm) and the presence of same-language subtitles (present vs. absent).

In the first part of Study 1, we investigated whether a narrator's emotional stance, conveyed solely through their voice, would affect participants' learning about sustainable construction from a video with a non-onscreen instructor. Two types of instructors were used: one enthusiastic and one calm. We formulated six hypotheses based on previous research and the cognitive-affective model of e-learning (Mayer, 2020). Participants rated the enthusiastic narrator higher in enthusiasm and activation, and the calm narrator higher in calmness and boredom, supporting Hypothesis 1. The enthusiastic narrator was perceived as more facilitative, credible, engaging, and human-like, supporting Hypothesis 2. However, no significant differences were found in positive activation, negative activation, or valence, contradicting Hypothesis 3. Similarly, no significant differences were observed in situational or delayed interest, intrinsic motivation, cognitive load, or mental effort, challenging Hypotheses 4 and 5. Regarding learning outcomes, there were no significant differences in knowledge, retention, transfer, certainty in answers, or self-evaluated test performance immediately and one week after the test. When accounting for potentially confounding variables, such as prior interest in the topic, prior knowledge, English proficiency, and initial emotional state, however, marginally significant differences were found in the transfer part of the test and self-evaluation of one's test performance, with those watching the calmer narrator achieving slightly higher results. While these results may not be enough to support Hypothesis 6, they might indicate a pattern that deserved further attention.

Additional analyses showed that participants with lower English proficiency who viewed videos with an enthusiastic narrator perceived the instructor more positively, reported higher intrinsic cognitive load, lower self-evaluated test performance, but performed better in both retention and transfer. In contrast, participants with higher proficiency performed worse in immediate and delayed tests but were more confident in their answers, indicating that the enthusiastic narrator hindered more proficient learners. This not only offered support for Hypothesis 6 but suggested that the narrator's emotional tone affects learners differently based on language proficiency.

Furthermore, another goal of Study 1 was also to verify whether embedding same-language subtitles into the videos may impact the cognitive load and learning outcomes among non-native English speakers, leading to the proposal of two additional hypotheses. Participants who watched videos with SLS reported slightly lower extraneous cognitive load compared to those without SLS (the difference was marginally significant), providing support for Hypothesis 7 and indicating that SLS may help reduce cognitive load by providing additional textual support to those learning academic content in a foreign language.

Although there were no differences in learning outcomes in the immediate knowledge test, marginally significant differences emerged in learning outcomes a week after the initial video viewing, especially in the transfer portion of the test, after adjusting for potential confounding variables. This provided partial support for Hypothesis 8 and suggests that SLS might positively affect more complex learning in the long run. Further analyses revealed that this effect was evident especially among learners with lower English proficiency, but not for those with better command of English, again highlighting a nuanced effect that warrants additional investigation.

Overall, previous research has demonstrated that an onscreen instructor exhibiting positive emotions can enhance learners' emotional states, cognitive processing, and learning outcomes (Lawson et al., 2021a, 2021b, 2021c). Our study further established that not only valence, but the activation level of an instructor's emotional expression as well can influence learners, even when conveyed solely through vocal pitch without additional social cues. This effect varies according to learners' language proficiency, underscoring the need for research focused on non-native learners and the inclusion of their individual differences. Adding SLS also emerged as a helpful tool for learners with lower proficiency in the language of the videos.

The aim of Study 2, on the other hand, was to determine how background music with different levels of activation affects the emotional, cognitive, and learning outcomes of students learning from videos in a foreign language. As part of the experiment, students were divided into three groups and watched the same educational videos as in Study 1 (with the enthusiastic narrator and without SLS) under different conditions: without added music (control group), with added calm music (experimental group 1), and with lively, uplifting music (experimental group 2). The experiment replicated the previous study: 307 university students, whether present in person or online, first completed a pre-test to assess their prior knowledge on the video topic, followed by an English proficiency test and a questionnaire to determine their emotional state. After watching the videos,

they filled out a post-questionnaire addressing various emotional and cognitive outcomes and then took a knowledge test. A week later, participants were invited to participate in the second phase, which involved retaking the same knowledge test they had completed seven days earlier.

Results show that participants from the calm music condition perceived the videos as more pleasant than those in the lively and no music conditions. While Hypothesis 9 predicted differences in positive activating emotions such as enthusiasm and excitement between groups, the results indicated that background music primarily affected negative activation. The calm music group experienced a greater reduction in negative activating emotions (e.g., nervousness, distress) compared to the lively music group. Although Hypothesis 9 was not confirmed, there is evidence that background music in learning videos influenced participants' emotional states, albeit in an unexpected manner. Instead of the lively music increasing excitement and energy, the calm music helped participants feel more relaxed and calm, indicating the importance of investigating the effect of music with varying energy levels. Those in the calm music group also reported a higher ability to focus on the lesson, enjoying the lesson more, and expressed more interest in the lesson, wanting more lessons like it. While there were differences in affective variables, these did not translate into cognitive processing, as no significant differences were found in cognitive load, leading us to reject Hypothesis 10. In contrast to Hypothesis 11, there were also no differences in objective learning performance, but the calm music group participants reported higher levels of self-evaluated test performance both immediately and a week after the learning episode.

The positive impact of adding calm background music to videos on participants' emotional state was even more pronounced among those with lower English proficiency. However, it appeared to negatively affect their mental effort and attention, especially compared to more energetic music. Conversely, videos with lively music seemed to demand more cognitive resources than no or calm music for participants with a good command of English, indicating that the effect of music varies based on language proficiency. The type of background music did not significantly influence students from study programs more familiar with the video content. However, adding calm background music had a beneficial effect on learners' emotional state, particularly for students less familiar with the video topic.

Similar to Study 1, the results of Study 2 highlight the importance of considering individual differences among participants. Findings from studies on native speakers cannot simply be applied to non-native speakers, as language proficiency appears to be

crucial for learning from videos, setting boundary conditions for multimedia learning principles. This dissertation contributes to guidelines for creating more effective educational videos by incorporating emotional design and focusing on an often overlooked group – non-native language learners. Additionally, while much of the multimedia learning literature emphasizes visual channel interventions, this dissertation advances understanding of the auditory channel, filling a significant gap in current knowledge. The main limitation of both studies is the small and non-representative sample, which is especially important when studying interventions where smaller effects are expected. Many of the findings are based on self-reporting, which is why future research would also benefit by including psychophysiological measurements that would provide more objective insight into students' emotional and cognitive processing. In the future, researchers should also explore different video content areas; our focus on sustainable construction may not have engaged participants fully. Testing various subject areas is essential, as the impact of multimedia learning interventions can vary depending on the instructional domain (Beege et al., 2023). By addressing these gaps, future studies can further refine educational video design and enhance learning outcomes for diverse populations.

5 REFERENCES

- Alapieti, T., Mikkola, R., Pasanen, P., & Salonen, H. (2020). The influence of wooden interior materials on indoor environment: a review. *European Journal of Wood and Wood Products*, 78, 617–634. <https://doi.org/10.1007/s00107-020-01532-x>
- Alley, T. R., & Greene, M. E. (2008). The relative and perceived impact of irrelevant speech, vocal music and non-vocal music on working memory. *Current Psychology*, 27(4), 277–289. <https://doi.org/10.1007/s12144-008-9040-z>
- Asdrubali, F., Ferracuti, B., Lombardi, L., Guattari, C., Evangelisti, L., & Grazieschi, G. (2017). A review of structural, thermo-physical, acoustical, and environmental properties of wooden materials for building applications. *Building and Environment*, 114, 307–332. <https://doi.org/10.1016/J.BUILDENV.2016.12.033>
- Aspin, D. N., & Chapman, J. D. (2000). Lifelong learning: Concepts and conceptions. *International Journal of Lifelong Education*, 19(1), 2–19. <https://doi.org/10.1080/026013700293421>
- Audacity Team. (2021). *Audacity®: Free Audio Editor and Recorder* (3.1.3). <https://audacityteam.org/>
- Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 8, pp. 47–89). Academic Press.
- Beege, M., & Schneider, S. (2023). Emotional design of pedagogical agents: The influence of enthusiasm and model-observer similarity. *Educational Technology Research and Development*, 71, 859–880. <https://doi.org/10.1007/s11423-023-10213-4>
- Beege, M., Schneider, S., Nebel, S., & Rey, G. D. (2020). Does the effect of enthusiasm in a pedagogical agent's voice depend on mental load in the learner's working memory? *Computers in Human Behavior*, 112, 106483. <https://doi.org/10.1016/j.chb.2020.106483>
- Beege, M., Schroeder, N. L., Heidig, S., Rey, G. D., & Schneider, S. (2023). The instructor presence effect and its moderators in instructional video: A series of meta-analyses. *Educational Research Review*, 41(November), 100564. <https://doi.org/10.1016/j.edurev.2023.100564>
- Bender, L., Renkl, A., & Eitel, A. (2021). Seductive details do their damage also in longer

- learning sessions – When the details are perceived as relevant. *Journal of Computer Assisted Learning*, 37(5), 1248–1262. <https://doi.org/10.1111/jcal.12560>
- Boersma, P., & Weenink, D. (2023). *Praat: doing phonetics by computer* (6.3.18). <http://www.praat.org/>
- Bonoli, A., Zanni, S., & Serrano-Bernardo, F. (2021). Sustainability in building and construction within the framework of circular cities and European New Green Deal. The contribution of concrete recycling. *Sustainability*, 13(4), 2139.
- 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>
- Brom, C., Stárková, T., & D'Mello, S. K. (2018). How effective is emotional design? A meta-analysis on facial anthropomorphisms and pleasant colors during multimedia learning. *Educational Research Review*, 25, 100–119. <https://doi.org/10.1016/j.edurev.2018.09.004>
- Bucik, V. (1997). *Osnove psihološkega testiranja*. Filozofska fakulteta Univerze v Ljubljani, Oddelek za psihologijo.
- Burnard, M., & Kutnar, A. (2015). Wood and human stress in the built indoor environment: a review. *Wood Science and Technology*, 49, 969–986. <https://doi.org/10.1007/s00226-015-0747-3>
- Cassidy, G., & MacDonald, R. A. R. (2007). The effect of background music and background noise on the task performance of introverts and extraverts. *Psychology of Music*, 35(3), 517–537. <https://doi.org/https://doi.org/10.1177/0305735607076444>
- Choi, H., & Johnson, S. D. (2005). The effect of context-based video instruction on learning and motivation in online courses. *American Journal of Distance Education*, 19(4), 215–227. https://doi.org/10.1207/s15389286ajde1904_3
- Christenson, S. L., Reschly, A. L., & Wylie, C. (2012). *Handbook of Research on Student Engagement*.
- Chung, S., Cheon, J., & Lee, K. W. (2015). Emotion and multimedia learning: An investigation of the effects of valence and arousal on different modalities in an instructional animation. *Instructional Science*, 43(5), 545–559. <https://doi.org/10.1007/s11251-015-9352-y>

- Clarke, L., & Sahin-Dikmen, M. (2020). Unions and the green transition in construction in Europe: Contrasting visions. *European Journal of Industrial Relations*, 26(4), 401–418. <https://doi.org/https://doi.org/10.1177/0959680120951705>
- Collins, M. L. (1978). Effects of enthusiasm training on preservice elementary teachers. *Journal of Teacher Education*, 29(1), 53–57. <https://doi.org/10.1177/002248717802900120>
- Colliot, T., & Jamet, É. (2018). Understanding the effects of a teacher video on learning from a multimedia document: An eye-tracking study. *Educational Technology Research and Development*, 66(6), 1415–1433. <https://doi.org/10.1007/s11423-018-9594-x>
- Colman, A. M. (2014). A Dictionary of Psychology. In *Oxford Reference* (3rd ed.). Oxford University Press.
- Corduban, C., Bochicchio, G., Polastri, A., & Ceccotti, A. (2012). Fire protection as a means to increase the sustainability of wood structures. *Bulletin UASVM Agriculture*, 69(2), 38–47. <https://doi.org/10.15835/BUASVMCN-AGR:8727>
- Davis, R. O., & Vincent, J. (2019). Sometimes more is better: Agent gestures, procedural knowledge and the foreign language learner. *British Journal of Educational Technology*, 50(6), 3252–3263. <https://doi.org/10.1111/bjet.12732>
- Davis, R. O., Vincent, J., & Park, T. J. (2019). Reconsidering the voice principle with non-native language speakers. *Computers and Education*, 140, 103605. <https://doi.org/10.1016/j.compedu.2019.103605>
- de la Mora Velasco, E., & Hirumi, A. (2020). The effects of background music on learning: A systematic review of literature to guide future research and practice. *Educational Technology Research and Development*, 68(6), 2817–2837. <https://doi.org/10.1007/s11423-020-09783-4>
- Delacre, M., Lakens, D., & Leys, C. (2017). Why psychologists should by default use Welch's t-test instead of Student's t-test. *International Review of Social Psychology*, 30(1), 92–101. <https://doi.org/10.5334/irsp.82>
- Dobbs, S., Furnham, A., & McClelland, A. (2011). The effect of background music and noise on the cognitive test performance of introverts and extraverts. *Applied Cognitive Psychology*, 313(25), 307–313. <https://doi.org/https://doi.org/10.1002/acp.1692>
- Du, M., Jiang, J., Li, Z., Man, D., & Jiang, C. (2020). The effects of background music on

- neural responses during reading comprehension. *Scientific Reports*, 10(1), 1–10. <https://doi.org/10.1038/s41598-020-75623-3>
- Echaide, C., Río, D., & Pacios, J. (2019). The differential effect of background music on memory for verbal and visuospatial information. *The Journal of General Psychology*, 146(4), 443–458. <https://doi.org/10.1080/00221309.2019.1602023>
- Ekkekakis, P. (2013). The Measurement of Affect, Mood, and Emotion: A Guide for Health-Behavioral Research. In *Cambridge University Press*.
- Endres, T., Weyreter, S., Renkl, A., & Eitel, A. (2020). When and why does emotional design foster learning? Evidence for situational interest as a mediator of increased persistence. *Journal of Computer Assisted Learning*, 36(4), 514–525. <https://doi.org/10.1111/jcal.12418>
- European Commission. (n.d. - a). *Construction*. Retrieved 4. 3. 2024, from https://single-market-economy.ec.europa.eu/sectors/construction_en
- European Commission. (n.d. - b). *New European Bauhaus*. Retrieved 7. 7. 2024, from https://new-european-bauhaus.europa.eu/index_en
- Eysenck, H. J. (1967). *The Biological Basis of Personality*. Thomas.
- Faculty of Social Sciences University of Ljubljana. (2022). 1KA (21.11.16). <https://www.1ka.si>
- Falk, B. (2009). Wood as a sustainable building material. *Forest Products Journal*, 59(9), 6–13. https://doi.org/10.5176/2301-394x_ace18.43
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175–191. <https://doi.org/https://doi.org/10.3758/BF03193146>
- Field, A. (2018). *Discovering Statistics using IBM SPSS Statistics* (5th ed.). SAGE Publications Ltd.
- Forest & Wood Products Australia. (2018). *The Environmental Advantages of Wood-Based Construction*.
- Gagnon, L., & Peretz, I. (2003). Mode and tempo relative contributions to “happy-sad” judgements in equitone melodies. *Cognition and Emotion*, 17(1), 25–40. <https://doi.org/10.1080/026999303022279>
- Gan, W., Chen, C., Wang, Z., Song, J., Kuang, Y., He, S., Mi, R., Sunderland, P., & Hu,

- L. (2019). Dense, self-formed char layer enables a fire-retardant wood structural material. *Advanced Functional Materials*, 29(20), 1902189. <https://doi.org/https://doi.org/10.1002/adfm.201807444>
- Gernsbacher, M. A. (2015). Video captions benefit everyone. *Policy Insights from the Behavioral and Brain Sciences*, 2(1), 195–202. <https://doi.org/https://doi.org/10.1177/2372732215602130>
- Goldhahn, C., Cabane, É., & Chanana, M. (2021). Sustainability in wood materials science: an opinion about current material development techniques and the end of lifetime perspectives. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 379, 20200339. <https://doi.org/10.1098/rsta.2020.0339>
- Green Building Council. (2020). *Policy Recommendations for Sustainable Building Materials*.
- Harp, S. F., & Mayer, R. E. (1997). The role of interest in learning from scientific text and illustrations: On the distinction between emotional interest and cognitive interest. *Journal of Educational Psychology*, 89(1), 92–102. <https://doi.org/10.1037/0022-0663.89.1.92>
- Harrell, F. E., Lee, K. L., Califf, R. M., Pryor, D. B., & Rosati, R. A. (1984). Regression modelling strategies for improved prognostic prediction. *Statistics in Medicine*, 3(2), 143–152. <https://doi.org/10.1002/sim.4780030207>
- Heidig, S., Beege, M., Rey, G. D., & Schneider, S. (2024). Instructor presence in instructional videos in higher education: three field experiments in university courses. *Educational Technology Research and Development*, 0123456789. <https://doi.org/10.1007/s11423-024-10391-9>
- Hill, C., & Dibdiakova, J. (2016). The environmental impact of wood compared to other building materials. *International Wood Products Journal*, 7(4), 215–219. <https://doi.org/10.1080/20426445.2016.1190166>
- Ho, H. Y., & Loo, F. Y. (2023). A theoretical paradigm proposal of music arousal and emotional valence interrelations with tempo, preference, familiarity, and presence of lyrics. *New Ideas in Psychology*, 71(March), 101033. <https://doi.org/10.1016/j.newideapsych.2023.101033>
- Horovitz, T., & Mayer, R. E. (2021). Learning with human and virtual instructors who display happy or bored emotions in video lectures. *Computers in Human Behavior*,

119, 106724. <https://doi.org/10.1016/j.chb.2021.106724>

Huberty, C. J., & Morris, J. D. (1989). Multivariate analysis versus multiple univariate analyses. *Psychological Bulletin*, 105(2), 302–308. <https://doi.org/10.1037/0033-2909.105.2.302>

Husain, G., Thompson, W. F., Schellenberg, E. G., Husain, G., Thompson, W. F., & Schellenberg, E. G. (2002). Effects of musical tempo and mode on arousal, mood, and spatial abilities. *Music Perception*, 20(2), 151–171. <https://doi.org/https://doi.org/10.1525/mp.2002.20.2.151>

IBM Corp. (2022). *IBM SPSS Statistics for Windows* (29.0.0.0). IBM Corp.

Isen, A. M., & Reeve, J. (2005). The influence of positive affect on intrinsic and extrinsic motivation: Facilitating enjoyment of play, responsible work behavior, and self-control. *Motivation and Emotion*, 29(4), 297–325. <https://doi.org/10.1007/s11031-006-9019-8>

Jäncke, L., & Sandmann, P. (2010). Music listening while you learn: No influence of background music on verbal learning. *Behavioral and Brain Functions*, 6(3), 1–14. <https://doi.org/10.1186/1744-9081-6-3>

Jiang, F., Li, T., Li, Y., Zhang, Y., Gong, A., Dai, J., Hitz, E., Luo, W., & Hu, L. (2018). Wood-based nanotechnologies toward sustainability. *Advanced Materials*, 30(1), 1703453. <https://doi.org/10.1002/adma.201703453>

John, O. P., Donahue, E. M., & Kentle, R. L. (1991). *The Big Five Inventory - Versions 4a and 54*. University of California, Berkeley, Institute of Personality and Social Research.

Jonsson, R., Rinaldi, F., Pilli, R., Fiorese, G., Hurmekoski, E., Cazzaniga, N., Robert, N., & Camia, A. (2021). Boosting the EU forest-based bioeconomy: Market, climate, and employment impacts. *Technological Forecasting and Social Change*, 163, 120478. <https://doi.org/https://doi.org/10.1016/j.techfore.2020.120478>

Juslin, P. N., & Laukka, P. (2003). Communication of Emotions in Vocal Expression and Music Performance: Different Channels, Same Code? *Psychological Bulletin*, 129(5), 770–814. <https://doi.org/10.1037/0033-2909.129.5.770>

Juslin, P. N., & Västfjäll, D. (2008). Emotional responses to music: The need to consider underlying mechanisms. *Behavioral and Brain Sciences*, 31(5), 559–621. <https://doi.org/10.1017/S0140525X08005293>

- Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist*, 38(1), 23–31. https://doi.org/10.1207/S15326985EP3801_4
- Kämpfe, J., Sedlmeier, P., & Renkewitz, F. (2010). The impact of background music on adult listeners: A meta-analysis. *Psychology of Music*, 39(4), 424–448. <https://doi.org/10.1177/0305735610376261>
- Kapwing. (2021). Kapwing©. [kapwing.com](https://www.kapwing.com)
- Killgore, W. D. S. (1998). The affect grid: A moderately valid, nonspecific measure of pleasure and arousal. *Psychological Reports*, 83(2), 639–642. <https://doi.org/10.2466/pr0.1998.83.2.639>
- Kind, T., & Evans, Y. N. (2015). Social media for lifelong learning. *International Review of Psychiatry*, 27(2), 124–132. <https://doi.org/10.3109/09540261.2014.990421>
- Klepsch, M., Schmitz, F., & Seufert, T. (2017). Development and validation of two instruments measuring intrinsic, extraneous, and germane cognitive load. *Frontiers in Psychology*, 8, 1–18. <https://doi.org/10.3389/fpsyg.2017.01997>
- Klepsch, M., & Seufert, T. (2020). Understanding instructional design effects by differentiated measurement of intrinsic, extraneous, and germane cognitive load. *Instructional Science*, 48(1), 45–77. <https://doi.org/10.1007/s11251-020-09502-9>
- Korbach, A., Brünken, R., & Park, B. (2017). Measurement of cognitive load in multimedia learning: A comparison of different objective measures. *Instructional Science*, 45(4), 515–536. <https://doi.org/10.1007/s11251-017-9413-5>
- Korbach, A., Brünken, R., & Park, B. (2018). Differentiating different types of cognitive load: A comparison of different measures. *Educational Psychology Review*, 30(2), 503–529. <https://doi.org/10.1007/s10648-017-9404-8>
- Kraft, M. A. (2020). Interpreting Effect Sizes of Education Interventions. *Educational Researcher*, 49(4), 241–253. <https://doi.org/10.3102/0013189X20912798>
- Krieglstein, F., Beege, M., Rey, G. D., Ginns, P., Krell, M., & Schneider, S. (2022). A systematic meta-analysis of the reliability and validity of subjective cognitive load questionnaires in experimental multimedia learning research. *Educational Psychology Review*, 34(4), 2485–2541. <https://doi.org/10.1007/s10648-022-09683-4>
- Krumm, I. R., Miles, M. C., Clay, A., Carlos II, W. G., & Adamson, R. (2021). Making

- effective educational videos for clinical teaching. *Chest*, 161(3), 764–772. <https://doi.org/10.1016/j.chest.2021.09.015>
- Lang, F. R., John, D., Lüdtke, O., Schupp, J., & Wagner, G. G. (2011). Short assessment of the Big Five: Robust across survey methods except telephone interviewing. *Behavior Research Methods*, 43(2), 548–567. <https://doi.org/10.3758/s13428-011-0066-z>
- Lawson, A. P., & Mayer, R. E. (2021). The power of voice to convey emotion in multimedia instructional messages. *International Journal of Artificial Intelligence in Education*, 32, 971–990. <https://doi.org/10.1007/s40593-021-00282-y>
- Lawson, A. P., & Mayer, R. E. (2022). Does the emotional stance of human and virtual instructors in instructional videos affect learning processes and outcomes? *Contemporary Educational Psychology*, 70, 102080. <https://doi.org/10.1016/j.cedpsych.2022.102080>
- Lawson, A. P., Mayer, R. E., Adamo-Villani, N., Benes, B., Lei, X., & Cheng, J. (2021a). The positivity principle: Do positive instructors improve learning from video lectures? *Educational Technology Research and Development*, 69, 3101–3129. <https://doi.org/https://doi.org/10.1007/s11423-021-10057-w>
- Lawson, A. P., Mayer, R. E., Adamo-Villani, N., Benes, B., Lei, X., & Cheng, J. (2021b). Recognizing the emotional state of human and virtual instructors. *Computers in Human Behavior*, 114, 106554. <https://doi.org/10.1016/j.chb.2020.106554>
- Lawson, A. P., Mayer, R. E., Adamo-Villani, N., Benes, B., Lei, X., & Cheng, J. (2021c). Do learners recognize and relate to the emotions displayed by virtual instructors? *International Journal of Artificial Intelligence in Education*, 31, 134–153. <https://doi.org/10.1007/s40593-021-00238-2>
- Leahy, W., & Sweller, J. (2011). Cognitive load theory, modality of presentation and the transient information effect. *Applied Cognitive Psychology*, 25, 943–951. <https://doi.org/10.1007/s11251-015-9362-9>
- Lebeničnik, M., Pitt, I., & Istenič Starčič, A. (2020). Optimal multimedia combination for students with dyslexia. *Metodološki Zvezki*, 17(2), 30–48. <https://doi.org/10.51936/myhk6939>
- Lee, H., & Mayer, R. E. (2018). Fostering learning from instructional video in a second language. *Applied Cognitive Psychology*, 32(5), 648–654. <https://doi.org/10.1002/acp.3436>

- Lehmann, J. A. M., Hamm, V., & Seufert, T. (2019). The influence of background music on learners with varying extraversion: Seductive detail or beneficial effect? *Applied Cognitive Psychology*, 33(1), 85–94. <https://doi.org/10.1002/acp.3509>
- Lehmann, J. A. M., & Seufert, T. (2017). The influence of background music on learning in the light of different theoretical perspectives and the role of working memory capacity. *Frontiers in Psychology*, 8, 1–11. <https://doi.org/10.3389/fpsyg.2017.01902>
- Lehmann, J. A. M., & Seufert, T. (2018). Can music foster learning - Effects of different text modalities on learning and information retrieval. *Frontiers in Psychology*, 8, 1–11. <https://doi.org/10.3389/fpsyg.2017.02305>
- Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of english. *Behavior Research Methods*, 44, 325–343. <https://doi.org/10.3758/s13428-011-0146-0>
- Li, W., Wang, F., Mayer, R. E., & Liu, H. (2019). Getting the point: Which kinds of gestures by pedagogical agents improve multimedia learning? *Journal of Educational Psychology*, 111(8), 1382–1395. <https://doi.org/10.1037/edu0000352>
- Liew, T. W., Mat Zin, N. A., & Sahari, N. (2017). Exploring the affective, motivational and cognitive effects of pedagogical agent enthusiasm in a multimedia learning environment. *Human-Centric Computing and Information Sciences*, 7(9). <https://doi.org/10.1186/s13673-017-0089-2>
- Liew, T. W., & Tan, S. M. (2016). The effects of positive and negative mood on cognition and motivation in multimedia learning environment. *Educational Technology and Society*, 19(2), 104–115.
- Liew, T. W., Tan, S. M., Tan, T. M., & Kew, S. N. (2020). Does speaker's voice enthusiasm affect social cue, cognitive load and transfer in multimedia learning? *Information and Learning Sciences*, 121(3–4), 117–135. <https://doi.org/10.1108/ILS-11-2019-0124>
- Lin, J. J. H., Lee, Y. H., Wang, D. Y., & Lin, S. S. J. (2016). Reading subtitles and taking enotes while learning scientific materials in a multimedia environment: Cognitive load perspectives on EFL students. *Educational Technology and Society*, 19(4), 47–58.
- Liu, Y., Jang, B. G., & Roy-Campbell, Z. (2018). Optimum input mode in the modality and redundancy principles for university ESL students' multimedia learning.

- Computers and Education*, 127(August 2017), 190–200.
<https://doi.org/10.1016/j.compedu.2018.08.025>
- Lynar, E., Cvejic, E., Schubert, E., & Vollmer-Conna, U. (2017). The joy of heartfelt music: An examination of emotional and physiological responses. *International Journal of Psychophysiology*, 120(October 2016), 118–125.
<https://doi.org/10.1016/j.ijpsycho.2017.07.012>
- Matthew, G. (2020). The effect of adding same-language subtitles to recorded lectures for non-native, english speakers in e-learning environments. *Research in Learning Technology*, 28, 1–16. <https://doi.org/10.25304/rlt.v28.2340>
- Mayer, R. E. (2014). The Cambridge Handbook of Multimedia Learning. In *Cambridge University Press* (2nd ed.). Cambridge University Press.
<https://doi.org/10.1017/CBO9781139547369>
- Mayer, R. E. (2020). Searching for the role of emotions in e-learning. *Learning and Instruction*, 70, 101213. <https://doi.org/10.1016/j.learninstruc.2019.05.010>
- Mayer, R. E. (2024). The past, present, and future of the cognitive theory of multimedia learning. *Educational Psychology Review*, 36(1), 1–25.
<https://doi.org/10.1007/s10648-023-09842-1>
- Mayer, R. E., & DaPra, C. S. (2012). An embodiment effect in computer-based learning with animated pedagogical agents. *Journal of Experimental Psychology: Applied*, 18(3), 239–252. <https://doi.org/10.1037/a0028616>
- Mayer, R. E., & Estrella, G. (2014). Benefits of emotional design in multimedia instruction. *Learning and Instruction*, 33, 12–18.
<https://doi.org/10.1016/j.learninstruc.2014.02.004>
- Mayer, R. E., & Fiorella, L. (2014). Principles for reducing extraneous processing in multimedia learning: Coherence, signaling, redundancy, spatial contiguity, and temporal contiguity principles. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 279–315). Cambridge University Press.
<https://doi.org/10.1017/CBO9781139547369.015>
- Mayer, R. E., Fiorella, L., & Stull, A. (2020). Five ways to increase the effectiveness of instructional video. *Educational Technology Research and Development*, 68(3), 837–852. <https://doi.org/10.1007/s11423-020-09749-6>
- Mayer, R. E., Lee, H., & Peebles, A. (2014). Multimedia learning in a second language: A cognitive load perspective. *Applied Cognitive Psychology*, 28(5), 653–660.

<https://doi.org/10.1002/acp.3050>

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

McNeish, D. (2017). Thanks coefficient alpha, we'll take it from here. *Psychological Methods*, 23(3), 412–433. <https://doi.org/10.1037/met0000144>

Merriam-Webster. (n.d. - a). *Background music*. Merriam-Webster.Com Dictionary. Retrieved 28. 3. 2023, from [https://www.merriam-webster.com/dictionary/background music](https://www.merriam-webster.com/dictionary/background%20music)

Merriam-Webster. (n.d. - b). *Incidental music*. Merriam-Webster.Com Dictionary. Retrieved 28. 3. 2023, from [https://www.merriam-webster.com/dictionary/incidental music](https://www.merriam-webster.com/dictionary/incidental%20music)

Meyerhoff, H. S., Merkt, M., Schröpel, C., & Meder, A. (2022). Medical education videos as a tool for rehearsal: Efficiency and the cases of background music and difficulty. *Instructional Science*, 50(6), 879–901. <https://doi.org/10.1007/s11251-022-09595-4>

Montero Perez, M., Van Den Noortgate, W., & Desmet, P. (2013). Captioned video for L2 listening and vocabulary learning: A meta-analysis. *System*, 41(3), 720–739. <https://doi.org/10.1016/j.system.2013.07.013>

Moreno, R. (2006). Does the modality principle hold for different media? A test of the method-affects-learning hypothesis. *Journal of Computer Assisted Learning*, 22(3), 149–158. <https://doi.org/10.1111/j.1365-2729.2006.00170.x>

Moreno, R., & Mayer, R. (2007). Interactive multimodal learning environments. *Educational Psychology Review*, 19(3), 309–326. <https://doi.org/10.1007/s10648-007-9047-2>

Moreno, R., & Mayer, R. E. (2000). A coherence effect in multimedia learning: The case for minimizing irrelevant sounds in the design of multimedia instructional messages. *Journal of Educational Psychology*, 92(1), 117–125. <https://doi.org/10.1037//0022>

Moussiades, L., Kazanidis, I., & Iliopoulou, A. (2019). A framework for the development of educational video: An empirical approach. *Innovations in Education and Teaching International*, 56, 217–228. <https://doi.org/10.1080/14703297.2017.1399809>

Navarro, D. J., & Foxcroft, D. R. (2022). *Learning statistics with jamovi: A tutorial for*

- psychology students and other beginners*.
<https://doi.org/https://doi.org/10.24384/hgc3-7p15>
- Paas, F. G. W. C. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology*, 84(4), 429–434. <https://doi.org/10.1037/0022-0663.84.4.429>
- Pannatier, M., & Béntrancourt, M. (2024). Learning from academic video with subtitles: When foreign language proficiency matters. *Learning and Instruction*, 90, 101863. <https://doi.org/10.1016/j.learninstruc.2023.101863>
- Pannatier, M., & Betrancourt, M. (2019). Effective learning from captioned video lecture in a foreign language: Literature review. *Educational Research Review*, 28, 100292. <https://doi.org/10.1016/j.edurev.2019.100292>
- Park, B., Flowerday, T., & Brünken, R. (2015). Cognitive and affective effects of seductive details in multimedia learning. *Computers in Human Behavior*, 44, 267–278. <https://doi.org/10.1016/j.chb.2014.10.061>
- Park, B., Plass, J. L., & Brünken, R. (2014). Cognitive and affective processes in multimedia learning. *Learning and Instruction*, 29, 125–127. <https://doi.org/10.1016/j.learninstruc.2013.05.005>
- Pekrun, R., & Linnenbrink-Garcia, L. (2014). *International Handbook of Emotions in Education*. Routledge. <https://doi.org/10.4324/9780203148211>
- Persson, J. R., Wattengård, E., & Lilledahl, M. B. (2019). The effect of captions and written text on viewing behavior in educational videos. *Lumat: International Journal of Math, Science and Technology Education*, 7(1), 124–147. <https://doi.org/10.31129/LUMAT.7.1.328>
- Peters, M., & Romero, M. (2019). Lifelong learning ecologies in online higher education: Students' engagement in the continuum between formal and informal learning. *British Journal of Educational Technology*, 50(4), 1729–1743. <https://doi.org/10.1111/BJET.12803>
- Plass, J. L., & Kalyuga, S. (2019). Four ways of considering emotion in cognitive load theory. *Educational Psychology Review*, 31(2), 339–359. <https://doi.org/10.1007/s10648-019-09473-5>
- Plass, J. L., & Kaplan, U. (2016). Emotional Design in Digital Media for Learning. In S. Y. Tettegah & M. Gartmeier (Eds.), *Emotions, Technology, Design, and Learning* (pp. 131–161). Elsevier Academic Press. <https://doi.org/10.1016/b978-0-12->

801856-9.00007-4

- Quade, D. (1979). Using weighted rankings in the analysis of complete blocks with additive block effects. *Journal of the American Statistical Association*, 74(367), 680–683. <https://doi.org/10.1080/01621459.1979.10481670>
- R Core Team. (2020). *R: A Language and environment for statistical computing* (4.0). <https://cran.r-project.org>
- Revelle, W., & Zinbarg, R. E. (2009). Coefficients alpha, beta, omega, and the glb: Comments on Sijtsma. *Psychometrika*, 74(1), 145–154. <https://doi.org/10.1007/s11336-008-9102-z>
- Rey, G. D. (2012). *A review of research and a meta-analysis of the seductive detail effect*. 7, 216–237. <https://doi.org/10.1016/j.edurev.2012.05.003>
- Rey, G. D., & Steib, N. (2013). The personalization effect in multimedia learning: The influence of dialect. *Computers in Human Behavior*, 29(5), 2022–2028. <https://doi.org/10.1016/j.chb.2013.04.003>
- Rotgans, J. I., & Schmidt, H. G. (2011). Situational interest and academic achievement in the active-learning classroom. *Learning and Instruction*, 21(1), 58–67. <https://doi.org/10.1016/j.learninstruc.2009.11.001>
- Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6), 1161–1178. <https://doi.org/10.1037/h0077714>
- Russell, J. A., Weiss, A., & Mendelsohn, G. A. (1989). Affect Grid: A Single-Item Scale of Pleasure and Arousal. *Journal of Personality and Social Psychology*, 57(3), 493–502. <https://doi.org/10.1037/0022-3514.57.3.493>
- Ruxton, G. D. (2006). The unequal variance t-test is an underused alternative to Student's t-test and the Mann-Whitney U test. *Behavioral Ecology*, 17(4), 688–690. <https://doi.org/10.1093/beheco/ark016>
- Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55(1), 68–78. <https://doi.org/10.1017/S1041610297004006>
- Ryu, J., & Baylor, A. L. (2005). The psychometric structure of Pedagogical Agent Persona. *Technology, Instruction, Cognition and Learning*, 2(4), 291–315.
- Sajinčič, N., Gordobil, O., Simmons, A., & Sandak, A. (2021). An exploratory study of consumers' knowledge and attitudes about lignin-based sunscreens and bio-based

- skincare products. *Cosmetics*, 8(3), 78.
<https://doi.org/https://doi.org/10.3390/cosmetics8030078>
- Salimpoor, V. N., Benovoy, M., Longo, G., Cooperstock, J. R., & Zatorre, R. J. (2009). The rewarding aspects of music listening are related to degree of emotional arousal. *PLoS ONE*, 4(10). <https://doi.org/10.1371/journal.pone.0007487>
- Salkind, N. (Ed.). (2010). *Encyclopedia of Research Design*. SAGE Publications Inc.
<https://doi.org/https://doi.org/10.4135/9781412961288>
- Sandak, A., Brzezicki, M., & Sandak, J. (2020). Trends and perspectives in the use of timber and derived products in building facades. In P. Samui, D. Kim, N. R. Iyer, & S. Chaudhary (Eds.), *New Materials in Civil Engineering* (pp. 333–374). Butterworth-Heinemann.
- Schallberger, U. (2005). Kurzskalen zur Erfassung der Positiven Aktivierung, Negativen Aktivierung und Valenz in Experience Sampling Studien (PANAVA-KS). [Short scales to assess positive activation, negative activation and valence in experience sampling studies (PANAVA-KS)]. In *Forschungsberichte aus dem Projekt: "Qualität des Erlebens in Arbeit und Freizeit"* (Issue 6).
- Schneider, S., Kriegelstein, F., Beege, M., & Daniel, G. (2022). The impact of video lecturers' nonverbal communication on learning – An experiment on gestures and facial expressions of pedagogical agents. *Computers and Education*, 176, 104350. <https://doi.org/10.1016/j.compedu.2021.104350>
- Schneider, S., Nebel, S., Pradel, S., & Rey, G. D. (2015). Introducing the familiarity mechanism: A unified explanatory approach for the personalization effect and the examination of youth slang in multimedia learning. *Computers in Human Behavior*, 43, 129–138. <https://doi.org/10.1016/j.chb.2014.10.052>
- Schneider, S., Wirzberger, M., & Rey, G. D. (2019). The moderating role of arousal on the seductive detail effect in a multimedia learning setting. *Applied Cognitive Psychology*, 33(1), 71–84. <https://doi.org/10.1002/acp.3473>
- Schroeder, N. L., Romine, W. L., & Craig, S. D. (2017). Measuring pedagogical agent persona and the influence of agent persona on learning. *Computers and Education*, 109, 176–186. <https://doi.org/10.1016/j.compedu.2017.02.015>
- Schroeder, N. L., Yang, F., Banerjee, T., Romine, W. L., & Craig, S. D. (2018). The influence of learners' perceptions of virtual humans on learning transfer. *Computers and Education*, 126, 170–182. <https://doi.org/10.1016/j.compedu.2018.07.005>

- Schubert, E. (2013). Emotion felt by the listener and expressed by the music: Literature review and theoretical perspectives. *Frontiers in Psychology*, 4(DEC), 1–18. <https://doi.org/10.3389/fpsyg.2013.00837>
- Shangguan, C., Wang, Z., Gong, S., Guo, Y., & Xu, S. (2020). More attractive or more interactive? The effects of multi-leveled emotional design on middle school students' multimedia learning. *Frontiers in Psychology*, 10, 3065. <https://doi.org/10.3389/fpsyg.2019.03065>
- Shuman, V., & Scherer, K. R. (2014). Concepts and structures of emotions. In P. A. Alexander, R. Pekrun, & L. Linnenbrink-Garcia (Eds.), *International Handbook of Emotions in Education* (pp. 13–35). Routledge. <https://doi.org/10.4324/9780203148211.ch2>
- Song, J., Chen, C., Zhu, S., Zhu, M., Dai, J., Ray, U., Li, Y., Kuang, Y., Li, Y., Quispe, N., Yao, Y., Gong, A., Leiste, U., Bruck, H. A., Zhu, J. Y., Vellore, A., Li, H., Minus, M. L., Jia, Z., ... Hu, L. (2018). Processing bulk natural wood into a high-performance structural material. *Nature*, 554, 224–228. <https://doi.org/10.1038/nature25476>
- Statistical Office of Republic of Slovenia. (2023a). *KLASIUS-P-16*. Retrieved 5. 5. 2023, from <https://www.stat.si/Klasius/Default.aspx?id=5>
- Statistical Office of Republic of Slovenia. (2023b). *KLASIUS-P-16 - Klasifikacija področij izobraževalnih aktivnosti/izidov 2016, V1*. Retrieved 5. 5. 2023, from <https://www.stat.si/Klasje/Klasje/Tabela/7541>
- Steffens, K. (2015). Competences, learning theories and MOOCs: Recent developments in lifelong learning. *European Journal of Education*, 50(1), 41–59. <https://doi.org/10.1111/EJED.12102>
- Su, Y., He, M., & Li, R. (2023). The effects of background music on English reading comprehension for English foreign language learners: Evidence from an eye movement study. *Frontiers in Psychology*, 14, 1–10. <https://doi.org/10.3389/fpsyg.2023.1140959>
- Sundararajan, N. K., & Adesope, O. (2020). Keep it coherent: A meta-analysis of the seductive details effect. *Educational Psychology Review*, 32(3), 707–734. <https://doi.org/10.1007/s10648-020-09522-4>
- Sung, E., & Mayer, R. E. (2012). When graphics improve liking but not learning from online lessons. *Computers in Human Behavior*, 28(5), 1618–1625.

<https://doi.org/10.1016/j.chb.2012.03.026>

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*, 4(4), 295–312.

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. Springer.

Sweller, J., van Merriënboer, J. J., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, 10(3), 251–296.

Szymkowiak, A., Melović, B., Dabić, M., Jeganathan, K., & Kundi, G. S. (2021). Information technology and Gen Z: The role of teachers, the internet, and technology in the education of young people. *Technology in Society*, 65, 101565. <https://doi.org/10.1016/J.TECHSOC.2021.101565>

Teigen, K. H. (1994). Yerkes-Dodson: A law for all seasons. *Theory & Psychology*, 4(4), 525–547.

Tellnes, L. G. F., Ganne-Chedeville, C., Dias, A., Dolezal, F., Hill, C., & Escamilla, E. Z. (2017). Comparative assessment for biogenic carbon accounting methods in carbon footprint of products: A review study for construction materials based on forest products. *IForest*, 10(5), 815–823. <https://doi.org/10.3832/ifor2386-010>

ten Hove, P., & van der Meij, H. (2015). Like it or not. What characterizes YouTube's more popular instructional videos? *Society for Technical Communication*, 62(1), 48–62.

The jamovi project. (2022). *jamovi* (1.6). <https://www.jamovi.org>

Thompson, W. F., Schellenberg, E. G., & Letnic, A. K. (2011). Fast and loud background music disrupts reading comprehension. *Psychology of Music*, 40(6), 700–708. <https://doi.org/10.1177/0305735611400173>

Tyng, C. M., Amin, H. U., Saad, M. N. M., & Malik, A. S. (2017). The influences of emotion on learning and memory. *Frontiers in Psychology*, 8, 1454. <https://doi.org/10.3389/fpsyg.2017.01454>

Um, E. R., Plass, J. L., Hayward, E. O., & Homer, B. D. (2012). Emotional design in multimedia learning. *Journal of Educational Psychology*, 104(2), 485–498. <https://doi.org/10.1037/a0026609>

United Nations. (2015). *Sustainable Development Goals*. Retrieved 1. 2. 2024, from <https://www.un.org/sustainabledevelopment/>

Van Breukelen, G. J. P. (2006). ANCOVA versus change from baseline had more power

- in randomized studies and more bias in nonrandomized studies. *Journal of Clinical Epidemiology*, 59(9), 920–925. <https://doi.org/10.1016/j.jclinepi.2006.02.007>
- van der Zee, T., Admiraal, W., Paas, F., Saab, N., & Giesbers, B. (2017). Effects of subtitles, complexity, and language proficiency on learning from online education videos. *Journal of Media Psychology*, 29(1), 18–30. <https://doi.org/10.1027/1864-1105/a000208>
- Vanlancker-Sidtis, D. (2003). Auditory recognition of idioms by native and nonnative speakers of English: It takes one to know one. *Applied Psycholinguistics*, 24(1), 45–57. <https://doi.org/10.1017/s0142716403000031>
- Watson, D., & Tellegen, A. (1985). Toward a consensual structure of mood. *Psychological Bulletin*, 98(2), 219–235. <https://doi.org/10.1037/0033-2909.98.2.219>
- Wilson, D., & Wharton, T. (2006). Relevance and prosody. *Journal of Pragmatics*, 38(10), 1559–1579. <https://doi.org/10.1016/j.pragma.2005.04.012>
- Wong, R. M., & Adesope, O. O. (2020). Meta-analysis of emotional designs in multimedia learning: a replication and extension study. *Educational Psychology Review*, 33, 1–29. <https://doi.org/10.1007/s10648-020-09545-x>
- WoodSolutions. (2019). *Technical Guide on Cross-Laminated Timber (CLT)*.
- Zhang, X., Lian, Z., & Ding, Q. (2016). Investigation variance in human psychological responses to wooden indoor environments. *Building and Environment*, 109, 58–67. <https://doi.org/10.1016/J.BUILDENV.2016.09.014>
- Zhang, X., Lian, Z., & Wu, Y. (2017). Human physiological responses to wooden indoor environment. *Physiology & Behavior*, 174, 27–34. <https://doi.org/10.1016/j.physbeh.2017.02.043>
- Zimmerman, D. W., & Zumbo, B. D. (1993). Rank transformations and the power of the Student t test and Welch t' test for non-normal populations with unequal variances. *Canadian Journal of Experimental Psychology / Revue Canadienne de Psychologie Expérimentale*, 47(3), 523–539. <https://doi.org/10.1037/h0078850>

SLUŠNO-ČUSTVENO OBLIKOVANJE PRI VEČPREDSTAVNOSTNEM UČENJU: IZOBRAŽEVALNI VIDEOPOSNETKI O LESU KOT GRADBENEM MATERIALU

KAZALO VSEBINE

1 UVOD.....	1
1.1 Les kot trajnosten in zdrav gradbeni material	1
1.2 Spletni izobraževalni videoposnetki kot orodje za vseživljenjsko učenje.....	3
2 TEORETIČNI DEL.....	6
2.1 Teorije učenja.....	6
2.1.1 Teorija kognitivne obremenitve	6
2.1.2 Kognitivna teorija večpredstavnostnega učenja	7
2.1.3 Vloga čustev pri učenju.....	9
2.1.4 Kognitivno-čustvena teorija učenja z mediji	11
2.1.5 Integrirani kognitivno-čustveni model učenja z multimedijo	13
2.1.6 Kognitivno-čustveni model e-učenja	13
2.2 Načela čustvenega oblikovanja	14
2.2.1 Minimalne manipulacije in zapeljive podrobnosti.....	15
2.2.2 Slušno-čustveno oblikovanje pri večpredstavnostnem učenju.....	17
2.2.2.1 Čustveni ton pripovedovalčevega glasu v izobraževalnih videih....	18
2.2.2.2 Glasba v izobraževalnih videih	20
2.3 Učenje v tujem jeziku z istojezičnimi podnapisi.....	23
3 EMPIRIČNI DEL.....	26
3.1 Raziskovalni problem, namen in cilji.....	26
3.2 Predštudija 1: Prepoznavanje človeških čustev iz pripovedovalčevega glasu	28
3.2.1 Raziskovalne hipoteze	28
3.2.2 Metodologija	28
3.2.2.1 Raziskovalni načrt.....	28
3.2.2.2 Udeleženci	28

3.2.2.3 Gradivo	30
3.2.2.4 Pripomočki	30
3.2.2.5 Postopek zbiranja podatkov	31
3.2.2.6 Postopek obdelave podatkov	31
3.2.3 Rezultati in razprava	31
3.3 Študija 1: Eksperiment o učinku čustvenega tona pripovedovalčevega glasu in istojezičnih podnapisih	34
3.3.1 Raziskovalne hipoteze	34
3.3.2 Metodologija	34
3.3.2.1 Raziskovalni načrt	34
3.3.2.2 Udeleženci	35
3.3.2.3 Gradivo	37
3.3.2.4 Pripomočki	39
3.3.2.5 Postopek zbiranja podatkov	48
3.3.2.6 Postopek obdelave podatkov	49
3.3.3 Rezultati in razprava	51
3.3.3.1 Opis in primerjava skupin	51
3.3.3.2 Čustveni ton pripovedovalca	57
3.3.3.3 Istojezični podnapisi	94
3.3.3.4 Interakcije	122
3.3.3.5 Omejitve in sklepi	130
3.4 Predštudiji 2: Ocenjevanje glasbe	132
3.4.1 Raziskovalne hipoteze	132
3.4.2 Metodologija	132
3.4.2.1 Raziskovalni načrt	132
3.4.2.2 Udeleženci	133
3.4.2.3 Gradivo	134
3.4.2.4 Pripomočki	135
3.4.2.5 Postopek zbiranja podatkov	136

3.4.2.6 Postopek obdelave podatkov	136
3.4.3 Rezultati in razprava	136
3.5 Študija 2: Eksperiment o učinku glasbe v ozadju	139
3.5.1 Raziskovalne hipoteze	139
3.5.2 Metodologija	139
3.5.2.1 Raziskovalni načrt	139
3.5.2.2 Udeleženci	139
3.5.2.3 Gradivo	143
3.5.2.4 Pripomočki	144
3.5.2.5 Postopek zbiranja podatkov	152
3.5.2.6 Postopek obdelave podatkov	154
3.5.3 Rezultati in razprava	156
3.5.3.1 Opis in primerjava skupin	156
3.5.3.2 Zaznava glasbe v ozadju in videoposnetkov	170
3.5.3.3 Čustveni izidi	175
3.5.3.4 Kognitivni izidi	194
3.5.3.5 Učni izidi	198
3.5.3.6 Dodatne analize	213
3.5.3.6 Omejitve in sklepi	223
4 SKLEPNE UGOTOVITVE	226
5 LITERATURA IN VIRI	231
SLUŠNO-ČUSTVENO OBLIKOVANJE PRI VEČPREDSTAVNOSTNEM UČENJU: IZOBRAŽEVALNI VIDEOPOSNETKI O LESU KOT GRADBENEM MATERIALU	248
KAZALO VSEBINE	248
POVZETEK VSEBINE IN UGOTOVITEV DOKTORSKE DISERTACIJE	252
6 PRILOGE	262
6.1 Priloga 1: Videoposnetki z njihovim trajanjem in pripadajočimi URL-ji	262
6.2 Priloga 2: Analiza zvočnih segmentov navdušenih in umirjenih pripovedi	264
6.3 Priloga 3: Korelacijska matrika (Pearsonov r) izidov Študije 1	268
6.4 Priloga 4: Pesmi, uporabljene v Predštudiji 2	270

6.5 Priloga 5: Predtestna vprašanja z odgovori (pravilni okrepljeni) v slovenščini, angleščini in norveščini	272
6.6 A Priloga 6: Testna vprašanja za priklic in transfer z odgovori (pravilni okrepljeni) v slovenščini, angleščini in norveščini	275
6.7 Priloga 7: Rezultati testa normalnosti in homogenosti za izide Študije 1	286
6.8 Priloga 8: Opisna statistika za glavne izide za udeležence z nizkim znanjem jezika glede na pripovedovalca	290
6.9 Priloga 9: ANCOVA primerjave za udeležence z nizkim znanjem jezika glede na pripovedovalca	294
6.10 Priloga 10: Opisna statistika za glavne izide za udeležence z visokim znanjem jezika glede na pripovedovalca	297
6.11 Priloga 11: ANCOVA primerjave za udeležence z visokim znanjem jezika glede na pripovedovalca.....	301
6.12 Priloga 12: Opisna statistika za glavne izide za udeležence z nizkim znanjem jezika glede na prisotnost podnapisov	304
6.13 Priloga 13: ANCOVA primerjave za udeležence z nizkim znanjem jezika glede na prisotnost podnapisov	308
6.14 Priloga 14: Opisna statistika za glavne izide za udeležence z visokim znanjem jezika glede na prisotnost podnapisov	311
6.15 Priloga 15: ANCOVA primerjave za udeležence z visokim znanjem jezika glede na prisotnost podnapisov	315
6.16 Priloga 16: Primerjave dvosmerne ANCOVA za skupino z nizkim znanjem jezika – Študija 1.....	318
6.17 Priloga 17: Primerjave dvosmerne ANCOVA za skupino z visokim znanjem jezika – Študija 1.....	322
6.18 Priloga 18: Seznam študijskih programov udeležencev Študije 2	326
6.19 Priloga 19: Korelacijska matrika (Pearsonov r) izidov Študije 2	327
6.20 Priloga 20: Rezultati testa normalnosti in homogenosti za izide Študije 2...	329
6.21 Priloga 21: Opisna statistika glede na znanje jezika – Študija 2	332
6.22 Priloga 22: ANCOVA primerjave za udeležence z nizkim znanjem jezika – Študija 2	341

6.23 ANCOVA primerjave za udeležence z visokim znanjem jezika – Študija 2	344
6.24 Priloga 24: Opisna statistika glede na poznavanje lesarstva – Študija 2....	347
6.25 Priloga 25: ANCOVA primerjave za lesarsko skupino – Študija 2	356
6.26 Priloga 26: ANCOVA primerjave za nelesarsko skupino – Študija 2	359

POVZETEK VSEBINE IN UGOTOVITEV DOKTORSKE DISERTACIJE

Trajnostni razvoj oziroma »zadovoljevanje potreb sedanjega človeškega rodu, ne da bi ogrožali možnosti prihodnjih rodov, da zadovoljijo svoje potrebe« (United Nations, 2015) je tema, ki zaradi svoje pomembnosti združuje znanost, politiko in industrijo. Zahteva sodelovanje ljudi po vsem svetu, zato je poznavanje tem trajnostnega razvoja splošna kompetenca, ki jo potrebuje vsak posameznik.

Eden glavnih ciljev trajnostnega razvoja je boj proti podnebnim spremembam (Združeni narodi, 2015). Javnost se zaveda pomembnosti uporabe obnovljivih materialov in virov energije, manj pa je poznano dejstvo, da je gradbena industrija odgovorna za približno 40 % emisij toplogrednih plinov in porabe energije v svetovnem merilu (Bonoli idr., 2021; Clarke in Sahin-Dikmen, 2020). Uporaba obnovljivih materialov v gradnji je zato vidik, ki ga je javnosti treba še posebej izpostaviti.

Les je naraven in obnovljiv vir, ki absorbira CO₂ in lahko nadomesti nekatere klasične, okolju škodljive gradbene materiale (Sandak idr., 2020). Čeprav ljudem gradnja z lesom ni tuja, je v zadnjih letih prišlo do številnih novosti, kot so modificiran les in inženirski lesni izdelki, ki ljudem ostajajo razmeroma neznani. Raziskave kažejo, da potrošniki trajnostne alternative poznamo zgolj površinsko, hkrati pa se poglobljeno znanje povezuje z večjo uporabo bolj trajnostnih možnosti (Sajinčič et al., 2021), zaradi česar je izobraževanje o gradnji z lesom pomembno tako za strokovno skupnost kot tudi za splošno javnost.

Priročen in učinkovit način za informiranje o trajnostnem razvoju in rešitvah proti podnebnim spremembam so izobraževalni videoposnetki. Z razvojem digitalne tehnologije in platform, kot je YouTube, sta dostopnost in uporaba izobraževalnih videoposnetkov močno narasli in pritegnili milijone učencev po vsem svetu. Videoposnetki spadajo med večpredstavnostna učna gradiva – gradiva, ki jih sestavljata pisna ali govorjena beseda in vizualne podobe, katerih cilj je spodbujanje učenja (Mayer,

2014). Glavni teoriji, na katerih temeljijo raziskave o načelih večpredstavnostnega učenja, sta kognitivna teorija večpredstavnostnega učenja (Mayer, 2014) in teorija kognitivne obremenitve (Sweller idr., 2011). Za razliko od kognitivnih, je vloga čustvenih (afektivnih), metakognitivnih in socialnih procesov manj raziskana. Raziskovalci so zato v zadnjih letih oblikovali celovitejše teorije učenja z večpredstavnostnimi gradivi, ki poleg kognitivnega delovanja vključujejo tudi čustva, motivacijo in metakognicijo.

Kognitivno-afektivna teorija učenja z mediji je razširila in nadgradila kognitivno teorijo večpredstavnostnega učenja, saj predpostavlja, da motivacija in čustva posredujejo v odnosu med kognitivnimi procesi in učenjem tako, da zmanjšajo ali povečajo količino kognitivnih virov, namenjenih učni nalogi (Moreno, 2006). V primerih, ko je na zaslonu prisoten tudi predavatelj, se predlaga kognitivno-čustveni model e-učenja (Mayer, 2020) z upoštevanjem petih korakov: predavatelj izraža čustveno stanje, študent prepozna čustveno stanje, razvije socialno povezavo s predavateljem, se zato pri učenju bolj potrudi, posledično pa se bolje odreže na testih znanja. Trenutno še ni odgovora, kako čustva, ki jih predavatelj izrazi, vplivajo na tiste, ki se učijo preko video vsebin in kako pri njih vzpodbuditi ustrezna čustva.

Pristop, ki skuša izboljšati izobraževalne videoposnetke z upravljanjem s čustveno-motivacijskimi dejavniki, je čustveno oblikovanje. Čustveno oblikovanje učnih gradiv vključuje dodajanje elementov z namenom spreminjanja čustev učencev oz. študentov na način, da le-ta spodbujajo učenje (Plass in Kaplan, 2016). Čustva lahko bistveno vplivajo na kognitivne procese, kot so zaznavanje, pozornost, učenje in spomin (Tyng idr., 2017), vendar se njihov učinek na učenje lahko razlikuje. Čustva imajo lahko blagodejno vlogo, saj posameznike preko zanimanja in veselja spodbudijo, da učenju namenijo več kognitivnih virov (Endres idr., 2020). Raziskave so na primer dokazale, da doživljanje pozitivnih čustev izboljšuje motivacijo in učenje (Um idr., 2012). Po drugi strani pa lahko čustva med učenjem predstavljajo tudi dodatno breme za procesiranje informacij, ker prinašajo dodatno kognitivno obremenitev, ki poslabša učne izide (Plass in Kalyuga, 2019).

Raziskave o načelih čustvenega oblikovanja lahko strnemo v dve veji: minimalne manipulacije in zapeljive podrobnosti. Prve so spremembe gradiva, namenjene spreminjanju motivacije in čustev, ne pa tudi učne vsebine. Uporaba okroglih, človeku podobnih oblik in toplih barv v predstavitvi, na primer, lahko sproža pozitivna čustva in olajša učenje (Wong in Adesope, 2020). Druga veja pa se nanaša na dodajanje zapeljivih podrobnosti ali informacij, ki so zanimive, vendar nepomembne za samo učno vsebino (Harp in Mayer, 1997), kot so okrasne slike, animacije, zabavna dejstva ali anekdote. V

nedavni metaanalizi se je izkazalo, da vključevanje zapeljivih podrobnosti prinaša mešane rezultate in da lahko tudi zavirajo učenje (Sundararajan in Adesope, 2020), ker tekmujejo za učenčevo pozornost. Večina raziskav s področja čustvenega oblikovanja se osredotoča na vizualne elemente, kot so oblike, barve in okrasne slike, veliko manj pa je znanega o vplivu slušnih vidikov večpredstavnostnih učnih gradiv.

Nedavna študija je pokazala, da študenti prepoznajo čustveni ton na podlagi glasu enako natančno kot ob prisotnosti osebe na zaslonu, ki ponuja dodatne neverbalne namige, kot so pogled, geste in telesna drža (Lawson in Mayer, 2021). Medtem ko lahko vključevanje posnetka predavatelja zahteva dodatne stroške, čas in delo, je pripoved neločljiv del večpredstavnostnega učnega gradiva, zato je spreminjanje čustvenega tona glasu minimalna manipulacija, ki jo je vredno dodatno preučiti. Raziskave kažejo, da smo ljudje posebej občutljivi na razlike med pozitivnim in negativnim čustvenim tonom v učnih gradivih (Lawson in Mayer, 2021) in da se ljudje bolje učimo, kadar predavatelji izražajo pozitivna čustva (Lawson idr., 2021a).

Čeprav so raziskave o vplivu čustev, ki jih izkazuje učitelj v videoposnetku, začele pridobivati na pozornosti, ostaja še mnogo odprtih vprašanj v zvezi z drugimi glasovnimi elementi v kontekstu izboljšanja učnega procesa. Poleg tega je dokazano, da pozitivna čustva pomagajo pri spodbujanju učenja, zato je smiselno, da le-ta podrobneje raziščemo, še posebej glede na njihovo stopnjo aktivacije, saj lahko slednja vpliva na učno uspešnost (Teigen, 1994). Afektivna (čustvena) stanja lahko opišemo kot kombinacijo dveh bipolarnih in ortogonalnih dimenzij, in sicer prijetnosti (valence) ter aktivacije oziroma stopnje fiziološke vzburjenosti (Russell, 1980; Watson in Tellegen, 1985). Navdušenje je na primer prijetno in aktivacijsko čustveno stanje, umirjenost pa prijetno, vendar deaktivacijsko stanje. Navdušenost učitelja se izraža preko ponavljajočih se neverbalnih vedenj, kot so vokalna animiranost, široko odprte oči, pogoste kretnje in gibi telesa ter visoka raven energije (Collins, 1978). Čustva v glasu pripovedovalca lahko izboljšajo motivacijo, čustvene in kognitivne izide študentov (Liew idr., 2017), hkrati pa lahko naredijo učno gradivo bolj zapleteno, kar poveča obremenitev pri obdelavi informacij in sčasoma negativno vpliva na učenje (Beege idr., 2020), zlasti pri učenju iz videoposnetka v tujem jeziku (Vanlancker-Sidtis, 2003).

Tako kot je pripovedovalčev glas primer minimalne manipulacije, je vključevanje glasbe v ozadju videoposnetka primer dodajanja zvočnih zapeljivih podrobnosti. Podobno kot čustveni ton glasu, ima tudi glasba v ozadju cilj vplivati na razpoloženje in fiziološko vzburjenost (Salimpoor idr., 2009), ki posledično delujeta na kognitivno procesiranje in učne dosežke (Husain idr., 2002). Fiziološko vzburjenje do neke mere

izboljšuje učenje, vendar lahko previsoka stopnja vzburjenja zmanjša učno uspešnost (Teigen, 1994). Vključevanje glasbe v izobraževalne videoposnetke se tako zdi preprost način za povečanje fiziološkega vzburjenja študentov, zlasti pesmi s hitrejšim tempom (Husain idr., 2002). Glasba v ozadju pa lahko predstavlja tudi dodatno kognitivno obremenitev, kar lahko negativno vpliva na učenje, zato njeni končni učinki niso vselej jasni in enoznačni.

Ugotovitve nedavne metaanalize (de la Mora Velasco in Hirumi, 2020) kažejo na pomanjkljivo proučevanje uporabe glasbe v izobraževalnih videoposnetkih; glasba je bila vključena v učna gradiva le v treh študijah, ki so pokazale pozitivne učinke na motivacijo, sposobnost priklica in učenje jezika. Na podlagi omenjenih študij se zdi, da se učinek glasbe v ozadju povečuje s trajanjem učenja. Temeljna študija, ki je obsegala dva eksperimenta o učinkih zvoka pri večpredstavnostnem učenju, ugotavlja, da je vključevanje glasbe v predstavitev poslabšalo učenje, vendar je bil uporabljeni zvočni posnetek dolg le 20 sekund in se je predvajal v zanki (Moreno in Mayer, 2000). Na splošno je v raziskavah o vplivu glasbe v ozadju na učenje prisotnih več metodoloških težav, zato je njihove ugotovitve težje primerjati in posploševati. Večina študij na primer ne poroča o uporabi glasbenih zvrsti ter stopnji aktivacije in valenci glasbe (de la Mora Velasco in Hirumi, 2020). Med tistimi, ki poročajo o uporabljenih skladbah, je najbolj priljubljena klasična glasba (Lehmann idr., 2019), zanemarjene pa so bolj sodobne zvrsti glasbe, kot je na primer ambientalna glasba, ki se najpogosteje uporablja kot glasbena podlaga v današnjih spletnih izobraževalnih videoposnetkih.

Druga raziskovalna vrzel v literaturi o čustvenem oblikovanju izobraževalnih videoposnetkov je ta, da večina raziskav temelji zgolj na učnih gradivih v maternem jeziku učencev oz. študentov. Zaradi razširjenosti in proste dostopnosti videoposnetkov v angleščini veliko ljudi posega po omenjenih gradivih, kar lahko povzroča dodatno breme pri učenju, kljub siceršnjemu poznavanju jezika. Izgovorjene besede so prehodne, kar lahko posameznikom s slabšim razumevanjem jezika predstavlja težavo pri obdelovanju govornih informacij (Leahy in Sweller, 2011). Po drugi stani pa so napisane besede na voljo dlje časa in omogočajo ponovni ogled oz. branje (Mayer idr., 2020). Podnapisi v jeziku videoposnetka so zato pogost način zagotavljanja dostopnosti gradiva širšemu občinstvu, vendar je lahko pisno besedilo tudi moteče, če skupaj z zvočno pripovedjo pretirano tekmuje za omejene kognitivne vire in s tem obremenjuje učni proces. Zato se zdi ključno preučevati tudi učne učinke podnapisov ob siceršnji zvočni pripovedi pri različnih populacijah.

Za osebe, ki se učijo v maternem jeziku, se je namreč izkazalo, da imajo lahko podnapisi v govorjenem jeziku škodljiv učinek na učenje (Lebeničnik idr., 2020; Mayer in Fiorella, 2014). Večina raziskav o učenju v tujem jeziku se je osredotočala na učinek podnapisov pri učenju zadevnega jezika in pokazala pozitivne rezultate (Perez idr., 2013), medtem ko raziskave o učenju vsebin v tujem jeziku, ki niso povezane z učenjem jezika, niso tako jasne, saj kažejo bodisi pozitivne (Lee in Mayer, 2018; Lin idr., 2016) bodisi nepomembne učinke na učenje (Liu idr., 2018; van der Zee idr., 2017).

Poleg strukturnih lastnosti izobraževalnega videoposnetka je pomembna tudi njegova učna vsebina. Študije večpredstavnostnega učenja namensko uvajajo učne vsebine, ki udeležencem niso znane (npr. Lehmann in Seufert, 2018; Liew in Tan, 2016). Predhodno znanje je namreč pomemben napovedni dejavnik pri oblikovanju video predavanj, saj ljudje z manjšim predhodnim znanjem obdelujejo informacije drugače kot študenti z več znanja, zato se optimalni načini predstavitve učnega gradiva med obema skupinama razlikujejo (Kalyuga idr., 2003). Z ustrezno izbiro teme in udeležencev, ki nimajo predznanja o dotični učni vsebini, omogočamo primerljivost raziskav, hkrati pa širimo spoznanja o vlogi dejavnikov oblikovanja posnetkov z različnih vsebinskih področij – v našem primeru gradnja z lesom. Preučevanje večpredstavnostnih učnih gradiv na to temo je še posebej pomembno, saj nobena dosedanja raziskava ni uporabila podobne učne vsebine; z raznolikostjo preučevanih učnih gradiv in vsebin pa omogočamo večjo posplošljivost raziskovalnih izsledkov.

Namen doktorske disertacije je bil raziskati, kako slušno-čustveno oblikovanje in podnapisi v govorjenem jeziku vplivajo na učenje iz videoposnetkov o lesu kot gradbenem materialu v tujem jeziku. Spoznanja smo pridobili na podlagi videoposnetkov na temo trajnostnega razvoja, ki je večini ljudem neznana, kljub njenemu bistvenemu prispevku v boju proti podnebnim spremembam. Za namene disertacije smo v sodelovanju s strokovnjaki za lesarstvo in gradbeništvo izdelali pet učnih videoposnetkov v angleščini z naslednjimi vsebinami: uvod v les kot trajnostni gradbeni material, procesi razgradnje lesa, premisleki o izbiri materiala, zaščitni ukrepi in prakse vzdrževanja.

Cilj disertacije je bil izvesti dva eksperimenta, s katerima smo raziskali učinke čustvenega tona pripovedovalca, izraženega le z glasom, vključevanja glasbe v ozadju z različnimi stopnjami aktivacije in podnapisov v govorjenem jeziku, ki se odražajo v učnih, kognitivnih in čustvenih spremenljivkah študentov. V naših raziskavah smo se osredotočili le na prijetna čustva, saj se v učnih videoposnetkih v praksi najpogosteje pojavljajo, pri čemer smo ločili med aktivacijskimi (navdušenje, živahna glasba) in

deaktivacijskimi (sproščenost, umirjena glasba) prijetnimi čustvi, ki so izražena bodisi z glasom pripovedovalke bodisi z glasbo v ozadju.

Cilj prve študije je bil preveriti, kako čustveni ton pripovedovalkega glasu in prisotnost podnapisov v govorjenem jeziku vplivata na učenje iz videoposnetka v tujem jeziku. Zanimalo nas je tudi, ali na morebitne razlike vpliva znanje angleščine udeležencev. Uporabili smo kvantitativno eksperimentalno raziskavo z deskriptivno in vzročno eksperimentalno metodo. Študija je bila izvedena kot 2 (čustveni ton pripovedi: miren in navdušen) x 2 (podnapisi: prisotni in odsotni) eksperiment z več skupinami in odvisnimi spremenljivkami: učni rezultati, kognitivna obremenitev in čustveno stanje.

Pred glavnim eksperimentom je bila na vzorcu 209 anketirancev izvedena tudi predštudija 1, s katero smo potrdili, da se navdušen in umirjen način pripovedovanja pomembno razlikujeta v stopnji aktiviranosti in lahko posnetke uporabimo kot neodvisno spremenljivko v glavnem eksperimentu.

V Študiji 1 je sodelovalo 229 študentov iz Slovenije in Norveške. Prevedli in priredili smo vprašalnike, ki se pogosto uporabljajo pri raziskovanju večpredstavnostnega učenja ter sestavili predhodni preizkus znanja ter preizkus znanja z vprašanji z več izbirami na temo lesa kot gradbenega materiala, ki so ga udeleženci rešili po ogledu videoposnetkov. Preizkus znanja je vseboval vprašanja, ki preverjajo tako zapornitev (retencija) kot prenos znanja (transfer). Uporabili smo tudi standardizirani test znanja angleščine. Teden dni po eksperimentu so bili študenti povabljeni k sodelovanju v drugem delu eksperimenta, v katerem so ponovno rešili preizkus znanja.

Rezultati, vezani na čustveni ton pripovedovalke so pokazali, da so udeleženci videoposnetke z navdušenim tonom pripovedovalkega glasu ocenili višje na lestvici navdušenja in nivoja aktivacije, videoposnetke z umirjenim tonom pripovedovanja pa višje na lestvici umirjenosti in dolgočasja, kar potrjuje hipotezo 1. Navdušen način pripovedovanja so zaznali kot bolj spodbuden, verodostojen in zanimiv, kar potrjuje hipotezo 2. Po drugi strani pa med skupinama ni bilo pomembnih razlik v čustvenem stanju udeležencev, torej v nivoju njihovih pozitivnih aktivacijskih čustev, negativnih aktivacijskih čustev ali valenci, kar ne potrjuje hipoteze 3. Podobno nismo našli pomembnih razlik v stopnji zanimanja za vsebino, notranje motivacije, kognitivne obremenitve ali miselnega napora udeležencev, kar se ne sklada s hipotezama 4 in 5, s katerima smo predvideli razlike v zanimanju in kognitivni obremenitvi. Poleg tega se je izkazalo, da ni bilo pomembnih razlik v znanju, retenciji, učnem transferu, stopnji prepričanosti v odgovore ali samooceni uspešnosti na preizkusu znanja takoj po eksperimentu in teden dni po njem. Ko smo pri analizi podatkov upoštevali tudi

spremenljivke, kot so predhodno zanimanje za temo, predhodno znanje, znanje angleščine in začetno čustveno stanje, pa so bile ugotovljene mejno pomembne razlike v delu testa, ki se nanaša na učni transfer in v samooceni svoje uspešnosti na preizkusu znanja, pri čemer so tisti, ki so spremljali videoposnetke z uporabo umirjenega tona glasu, dosegli nekoliko boljše rezultate. Čeprav na podlagi teh rezultatov nismo uspeli potrditi hipoteze 6, smo opravili dodatne analize in preverili rezultate ločeno za udeležence z nižjim in višjim nivojem znanja angleščine. Dodatne analize so pokazale, da so udeleženci s slabšim znanjem angleščine, ki so spremljali videoposnetke z navdušenim tonom pripovedovanja, so pripovedovalko dojemali bolj pozitivno, poročali so o večji notranji kognitivni obremenitvi in slabši samooceni uspešnosti na preizkusu znanja, vendar so imeli višje rezultate na preizkusu znanja, tako pri retenciji kot pri prenosu znanja. Nasprotno so se udeleženci z višjim znanjem angleščine in navdušenim načinom pripovedovanja slabše odrezali, tako na takojšnjem testu znanja kot na testu znanja teden dni po ogledu videoposnetkov, vendar so bili bolj prepričani v ustreznost svojih odgovorov. Slednje kaže na to, da je omenjeni način pripovedovanja oviral študente z višjim nivojem angleščine ter pomagal študentom z nižjim nivojem angleščine. Te ugotovitve niso le potrdile hipoteze 6, temveč pokazale na dejstvo, da je vpliv čustvenega tona glasu pripovedovalca odvisno od nivoja znanja jezika v videoposnetkih.

Poleg tega je bil cilj Študije 1 tudi preveriti, ali lahko podnapisi v govorjenem jeziku (v našem primeru v angleščini) vplivajo na kognitivno obremenitev in učne dosežke pri študentih, ki se učijo iz videoposnetkov v jeziku, ki jim ni materni. Udeleženci, ki so gledali videoposnetke s podnapisi, so poročali o nekoliko manjši zunanji kognitivni obremenitvi v primerjavi z udeleženci, ki so si ogledali videoposnetke brez podnapisov, pri čemer je bila razlika mejno pomembna. Kljub majhnemu učinku potrjujemo hipotezo 7, saj podatki nakazujejo na trend, da lahko podnapisi pomagajo zmanjšati kognitivno obremenitev pri učenju vsebine v tujem jeziku tako, da študentom ponudijo dodatno besedilno podporo.

Čeprav pri takojšnjem preizkusu znanja ni bilo razlik v učnih rezultatih med študenti, ki so si ogledali videoposnetke z ali brez podnapisov, so se po vključitvi morebitnih motečih spremenljivk v analize pojavile mejno pomembne razlike v učnih rezultatih teden dni po ogledu videoposnetkov, zlasti v delu preizkusa, ki se je nanašal na učni transfer. Ti rezultati delno potrjujejo hipotezo 8 in nakazujejo, da lahko podnapisi dolgoročno pozitivno vplivajo na kompleksnejše učenje. Nadaljnje analize so pokazale, da je bil ta učinek očiten zlasti pri študentih z nižjo ravno znanja angleščine, ne pa tudi pri tistih z višjo ravno znanja angleščine, kar ponovno kaže na obstoj specifičnih pogojev pri raziskovanju načel večpredstavnostnega učenja in potrebo po nadaljnjem preučevanju.

Če povzamemo, prejšnje raziskave so pokazale, da lahko predavatelj na zaslonu, ki izraža pozitivna čustva, izboljša čustveno stanje učencev, zmanjša njihovo kognitivno obremenitev in pozitivno vpliva na njihove učne rezultate (Lawson idr., 2021a, 2021b, 2021c). V naši študiji smo ugotovili, da ne le valenca, temveč tudi stopnja aktivacije čustvenega izražanja lahko vpliva na študente, tudi v primeru, če predavatelja ni na zaslonu in je čustveni ton izražen le z glasom, brez dodatnih neverbalnih znakov. Ta učinek se razlikuje glede na stopnjo znanja jezika študentov, kar nakazuje na to, da načel oblikovanja učnih gradiv ni mogoče neposredno posplošiti na populacijo ljudi, ki se učijo v tujem jeziku, zaradi česar je v prihodnjih raziskavah potrebno upoštevati tudi specifične individualne razlike med študenti. Zaključimo lahko tudi, da se je za študente s slabšim poznavanjem jezika, v katerem so se učili, prisotnost podnapisov v jeziku posnetka izkazalo kot koristno orodje, ki lahko nudi dodatno oporo pri učenju, zaradi česar je videoposnetkom vredno dodati tudi prepis govornjene vsebine.

Cilj Študije 2 je bil preučiti učinke glasbe v ozadju z različnimi stopnjami aktivacije na učne, kognitivne in čustvene rezultate študentov, ki se učijo novo snov v tujem jeziku. V okviru eksperimenta so bili študenti razdeljeni v tri skupine in si ogledali iste izobraževalne videoposnetke kot v Študiji 1 (z navdušenim tonom pripovedovanja in brez podnapisov) pod različnimi pogoji: brez glasbe v ozadju (kontrolna skupina), z dodano umirjeno glasbo (eksperimentalna skupina 1) in z živahno, spodbudno glasbo (eksperimentalna skupina 2). Skladbi, uporabljeni v glavnem eksperimentu, sta bili izbrani in validirani v dveh predštudijah. V prvi je 43 udeležencev ocenjevalo čustveni ton in raven energije 20 skladb. Skladbi z najvišjo in najnižjo ravno energije ter podobno oceno čustvenega tona sta bili izbrani in uporabljeni v naslednji predštudiji, v kateri je 66 udeležencev ocenilo svoje počutje ob obeh izbranih skladbah z vidika valence in ravni aktivacije, s čimer smo potrdili, da obe skladbi vplivata na čustveno stanje udeležencev na pričakovan način in ju lahko uporabimo v glavnem eksperimentu.

Študija 2 je potekala na enak način kot Študija 1. V eksperimentu je sodelovalo 307 univerzitetnih študentov, katerih naloga je bila najprej izpolniti test znanja za oceno nivoja predhodnega znanja na temo videoposnetka, opraviti test znanja angleščine in izpolniti vprašalnik o izhodiščnem čustvenem stanju. Po ogledu videoposnetkov so udeleženci izpolnili sklop vprašalnikov o različnih vidikih čustvenega in kognitivnega delovanja, nato pa so opravili še preizkus znanja o vsebini videoposnetkov. Teden dni pozneje so bili udeleženci povabljeni k sodelovanju v drugi fazi eksperimenta, ki je vključevala ponovno opravljanje istega preizkusa znanja, ki so ga opravili sedem dni prej.

Rezultati so pokazali, da so udeleženci, ki so si ogledali videoposnetke z umirjeno glasbo, le-te dojemali kot prijetnejše kot udeleženci, ki so gledali videoposnetke z živahno glasbo ali brez glasbe. Čeprav je hipoteza 9 predpostavljala, da se bodo skupine razlikovale v nivoju pozitivnih aktivacijskih čustev, kot sta navdušenje in vznemirjenje, se je izkazalo, da je glasba v ozadju vplivala predvsem na ravni čustev negativne aktivacije. Analize so namreč pokazale, da se je v skupini z umirjeno glasbo v primerjavi s skupino z živahno glasbo zmanjšala raven negativnih aktivacijskih čustev, kot sta npr. živčnost ali anksioznost. Čeprav hipoteza 9 ni bila potrjena, naši rezultati nakazujejo, da glasba v ozadju učnih videoposnetkov vpliva na čustveno stanje udeležencev, vendar v nepričakovani smeri. Namesto, da bi živahna glasba povečala vznemirjenje in energijo, je umirjena glasba udeležencem pomagala, da so se počutili bolj sproščeno in umirjeno, kar kaže na pomen raziskovanja učinka glasbe z različnimi ravnmi energije. Udeleženci iz skupine z umirjeno glasbo so poročali tudi o večji sposobnosti osredotočanja na vsebino, pri učenju so bolj uživali, izrazili so večje zanimanje za učno vsebino in si bolj želeli podobnih učnih gradiv. Kljub temu, da so se pokazale razlike v spremenljivkah čustvene narave, pa ni prišlo do pomembnih razlik v nivoju kognitivne obremenitve, zaradi česar nismo potrdili hipoteze 10. Glede na hipotezo 11 prav tako ni bilo razlik v objektivni učni uspešnosti, vendar so bili udeleženci skupine z umirjeno glasbo bolj samozavestni glede svoje uspešnosti na testu znanja takoj in teden dni po prvem delu eksperimenta kot udeleženci, ki so poslušali živahno glasbo.

Blagodejni učinek umirjene glasbe v ozadju na čustveno stanje je bil še izrazitejši pri udeležencih z nižjim znanjem angleščine, vendar pa se je pokazalo, da ta glasba negativno vpliva na stopnjo miselnega napora in usmerjanje pozornosti, zlasti v primerjavi z bolj energično glasbo. Po drugi strani so videoposnetki z živahno glasbo pri študentih z višjim znanjem angleščine zahtevali več kognitivnih virov kot videoposnetki brez glasbe ali z umirjeno glasbo, kar kaže, da se učinek glasbe razlikuje glede na stopnjo znanja jezika. Vrsta glasbe v ozadju ni bistveno vplivala na študente iz študijskih programov, ki so bolj povezani z vsebino videoposnetkov. Pri študentih iz programov, ki so manj povezani z učno tematiko in so manj seznanjeni s temo, pa je umirjena glasba v ozadju ugodno vplivala na njihovo čustveno stanje.

Podobno kot v Študiji 1, tudi rezultati Študije 2 poudarjajo pomen upoštevanja individualnih razlik med udeleženci. Disertacija prispeva k razumevanju, da ugotovitev iz študij, opravljenih na vzorcu študentov, ki se učijo snov v svojem maternem jeziku, ni mogoče preprosto prenesti na celotno populacijo, saj se je izkazalo, da je stopnja znanja jezika, v katerem je učna vsebina predstavljena, ključnega pomena. S tem smo odkrili in

izpostavili nov specifični pogoj, ki ga je potrebno upoštevati pri raziskovanju večpredstavnostnega učenja. Naše ugotovitve prispevajo k oblikovanju smernic za ustvarjanje učinkovitejših izobraževalnih videoposnetkov z vključevanjem čustvenega oblikovanja in osredotočanjem na pogosto spregledano skupino – ljudi, ki se učne snovi učijo v tujem jeziku. Poleg tega se literatura o večpredstavnostnem učenju osredotoča predvsem na oblikovna načela, ki zadevajo vidno zaznavanje, disertacija pa prispeva k razumevanju načel oblikovanja, ki podpirajo slušno zaznavanje, kar zapolnjuje pomembno vrzel v dosedanjem razumevanju oblikovanja večpredstavnostnih gradiv.

Glavna pomanjkljivost obeh študij je manjši in nereprezentativni raziskovalni vzorec, na kar je še posebej pomembno opozoriti pri preučevanju intervencij, pri katerih pričakujemo manjše učinke. Prav tako veliko ugotovitev temelji na samoporočanju, zaradi česar bi v prihodnje raziskave kazalo vključiti tudi psihofiziološke meritve, ki bi nudile bolj objektivne informacije o čustvenem in kognitivnem delovanju študentov. Smiselno bi bilo tudi, da nadaljnje raziskave uporabijo učna gradiva z različnih vsebinskih področij; tema naših videoposnetkov morda udeležencev ni dovolj pritegnila. Preučevanje različnih vsebinskih področij je bistvenega pomena, saj se je izkazalo, da se lahko vpliv načel večpredstavnostnega učenja razlikuje glede na učni predmet in vsebino (Beege idr., 2023). Z odpravljanjem teh vrzeli lahko prihodnje študije dodatno izboljšajo zasnovo učnih videoposnetkov in izboljšajo učne dosežke večjega števila ljudi.

6 APPENDIX

6.1 Appendix 1: Videos with their duration and corresponding URLs

Video clip	Duration	URL
Pre-study 1		
Enthusiastic 1	0:33	https://www.youtube.com/watch?v=GXPDbhGFW_U
Enthusiastic 2	0:46	https://www.youtube.com/watch?v=4fefLv_UlaQ
Enthusiastic 3	1:01	https://www.youtube.com/watch?v=upu3azd_UXs
Enthusiastic 4	0:37	https://www.youtube.com/watch?v=2cZJ28Et6SU
Enthusiastic 5	0:49	https://www.youtube.com/watch?v=zseQKBDCCpk
Calm 1	0:36	https://www.youtube.com/watch?v=TcQaByXR5xs
Calm 2	0:47	https://www.youtube.com/watch?v=0-mjNwVygS0
Calm 3	1:04	https://www.youtube.com/watch?v=G5wMaWZbEW8
Calm 4	0:41	https://www.youtube.com/watch?v=5D9OYjnzt3s
Calm 5	0:51	https://www.youtube.com/watch?v=MTsopcGWzbM
Study 1		
Enthusiastic 1	3:12	https://www.youtube.com/watch?v=lnGoPb1-0C8&t
Enthusiastic 2	5:48	https://www.youtube.com/watch?v=YLZHEMbtX58
Enthusiastic 3	6:37	https://www.youtube.com/watch?v=gEcw7By5xNg&t
Enthusiastic 4	3:48	https://www.youtube.com/watch?v=YDqwlYvRnKw&t
Enthusiastic 5	4:29	https://www.youtube.com/watch?v=GNalwW9rD38&t
Calm 1	3:12	https://www.youtube.com/watch?v=0_E821ff3nM
Calm 2	5:48	https://www.youtube.com/watch?v=taQXU71rjY0
Calm 3	6:43	https://www.youtube.com/watch?v=zwMrNc9YTKw
Calm 4	3:48	https://www.youtube.com/watch?v=x5XWFD6HT3c
Calm 5	4:36	https://www.youtube.com/watch?v=ilr6JR8ffkg
Enthusiastic & SLS 1	3:12	https://www.youtube.com/watch?v=00zucGZzoNo&t
Enthusiastic & SLS 2	5:48	https://www.youtube.com/watch?v=uAGHB0kbs_0
Enthusiastic & SLS 3	6:37	https://www.youtube.com/watch?v=tDCIsnirnpk
Enthusiastic & SLS 4	3:48	https://www.youtube.com/watch?v=uWGqcfGalNk
Enthusiastic & SLS 5	4:29	https://www.youtube.com/watch?v=rcS4_oi1Kqc
Calm & SLS 1	3:12	https://www.youtube.com/watch?v=UVGfQvNXZjo

Calm & SLS 2	5:48	https://www.youtube.com/watch?v=4zKGjR8PCs
Calm & SLS 3	6:43	https://www.youtube.com/watch?v=hsB-nCLEJLs
Calm & SLS 4	3:48	https://www.youtube.com/watch?v=hIWWNMBB_9A
Calm & SLS 5	4:36	https://www.youtube.com/watch?v=xTyTITMXpmc&t
Study 2		
No music 1	3:12	https://www.youtube.com/watch?v=InGoPb1-0C8&t
No music 2	5:48	https://www.youtube.com/watch?v=YLZHEMbtX58
No music 3	6:37	https://www.youtube.com/watch?v=gEcw7By5xNg&t
No music 4	3:48	https://www.youtube.com/watch?v=YDqwlYvRnKw&t
No music 5	4:29	https://www.youtube.com/watch?v=GNalwW9rD38&t
Calm music 1	3:12	https://www.youtube.com/watch?v=ixqguJ6s57o&t
Calm music 2	5:48	https://www.youtube.com/watch?v=YLYI8-E4SN8&t
Calm music 3	6:37	https://www.youtube.com/watch?v=n9VctME0Y3g&t
Calm music 4	3:48	https://www.youtube.com/watch?v=W7p3NI93fu4&t
Calm music 5	4:29	https://www.youtube.com/watch?v=6Qa3Y0grXD8&t
Lively music 1	3:12	https://www.youtube.com/watch?v=ysyz-7nNVpg&t
Lively music 2	5:48	https://www.youtube.com/watch?v=cGj8rpqFm8Y&t
Lively music 3	6:37	https://www.youtube.com/watch?v=leb2mUdBZV4&t
Lively music 4	3:48	https://www.youtube.com/watch?v=oM2yRFMTVQQ&t
Lively music 5	4:29	https://www.youtube.com/watch?v=Lfzjh41ItLs&t

6.2 Appendix 2: Pitch analysis of audio segments taken from the videos with enthusiastic and calm narrations

	Audio length (s)	Mdn (Hz)	M (Hz)	SD (Hz)	Min – Max (Hz)
Video 1					
<i>“Why wood? The construction industry consumes more resources than any other human activity.”</i>					
Enthusiastic	7.933	236.338	241.667	52.499	86.946–496.082
Calm	7.814	207.265	205.130	36.753	74.954–304.692
<i>“It is responsible for 40% of global energy use, raw material consumption and waste, and more than a third of CO₂ emissions.”</i>					
Enthusiastic	9.620	253.467	250.952	44.134	94.871–465.767
Calm	8.878	211.357	213.326	41.729	81.068–459.910
<i>“This means that the structure meets or exceeds functional performance requirements. But these are not the only requirements for a building.”</i>					
Enthusiastic	9.390	233.880	250.729	82.202	81.939–500.142
Calm	8.658	196.917	193.999	50.434	71.657–494.201
<i>“While a building can still be safe long after it has ceased to be functional, changes in the appearance of the material become noticeable much earlier...”</i>					
Enthusiastic	8.352	238.609	244.613	75.163	74.104–500.486
Calm	8.891	194.322	191.836	57.284	73.094–506.736
<i>“...and may become unacceptable long before the structure reaches the functional limit.”</i>					
Enthusiastic	4.258	219.894	209.997	56.087	81.674–287.644
Calm	4.365	182.152	166.094	45.565	76.980–478.183
Video 2					
<i>“Degradation processes. Without precautions, natural wood can deteriorate faster than other materials, ...”</i>					
Enthusiastic	7.556	240.396	244.087	62.779	83.389–483.593
Calm	8.353	193.820	191.132	38.484	78.010–499.758
<i>“... so knowing how to combat wood degradation can greatly extend its service life.”</i>					
Enthusiastic	5.088	231.358	236.360	61.205	83.354–368.757
Calm	4.994	177.627	165.529	58.937	79.893–478.639

<i>"There are three main degradation processes you should watch out for in relation to wooden structures: ..."</i>					
Enthusiastic	5.998	210.536	232.828	58.792	77.052–455.057
Calm	5.611	178.160	166.382	49.766	75.765–490.430
<i>"... weathering, rot caused by fungi, and insect activity."</i>					
Enthusiastic	5.610	232.734	245.017	71.674	85.743–489.692
Calm	4.599	181.173	165.105	43.393	75.798–220.273
<i>"With some initial planning and additional costs, we can choose better performing materials and design the structure for healthier, better-looking buildings ..."</i>					
Enthusiastic	8.500	237.938	245.650	61.913	79.417–475.417
Calm	9.368	199.148	201.598	47.668	78.878–490.962
<i>"... that save money and time used for regular maintenance."</i>					
Enthusiastic	3.433	215.123	209.829	73.453	79.959–486.333
Calm	3.388	195.071	193.002	44.715	78.317–465.035
Video 3					
<i>"Choosing the right material is not an easy task, as there are over 60 thousand wood species in the world."</i>					
Enthusiastic	6.370	236.580	253.318	62.785	84.170–395.551
Calm	6.275	191.006	190.373	43.043	70.002–309.796
<i>"Project materials are usually chosen based on initial investment and appearance, but other factors should also be considered, ..."</i>					
Enthusiastic	6.986	222.056	233.444	54.410	89.729–422.766
Calm	7.567	187.413	182.135	42.933	82.221–447.291
<i>"... including climate, maintenance needs and costs, performance over time, and compatibility with local culture and aesthetics."</i>					
Enthusiastic	9.093	223.638	235.760	61.905	82.801–487.331
Calm	8.617	199.799	191.089	36.727	75.508–269.057
<i>"Applications where there is little risk of rot don't need to use very durable wood, ..."</i>					
Enthusiastic	4.724	218.308	237.467	66.160	78.525–379.225
Calm	5.033	185.414	183.844	33.840	76.428–252.089
<i>"... but the higher the use class, the more durable species and greater protection are required."</i>					
Enthusiastic	5.438	206.715	221.273	57.491	79.631–428.856
Calm	5.587	183.729	181.754	38.581	80.521–466.108
Video 4					

<i>"After selecting appropriate materials, the simplest and most effective way to protect a wooden building is to design it ..."</i>					
Enthusiastic	6.781	226.928	239.535	60.552	82.456–470.213
Calm	7.476	193.420	194.121	33.190	76.615–460.744
<i>"... to eliminate or minimize negative influences on the material's natural drawbacks."</i>					
Enthusiastic	5.285	209.485	221.566	59.571	76.755–475.119
Calm	5.629	181.534	180.404	48.218	71.591–486.478
<i>"This approach can reduce future repairs, replacements and costs."</i>					
Enthusiastic	4.433	214.546	246.379	73.192	171.024–483.320
Calm	4.506	190.901	197.580	35.741	84.275–463.207
<i>"Last but not least, all engineered façade solutions should be designed to allow for easy maintenance and consist of individual elements ..."</i>					
Enthusiastic	8.511	226.773	244.687	69.466	73.922–483.735
Calm	9.302	192.318	191.101	37.737	76.178–483.539
<i>"...that can be disassembled and replaced without damaging the other structural elements."</i>					
Enthusiastic	4.682	219.595	227.310	44.989	148.643–479.888
Calm	5.119	186.090	189.314	53.106	81.196–491.410
Video 5					
<i>"Maintenance. Proper design can greatly reduce the rate of deterioration, but all materials need some kind of ongoing care."</i>					
Enthusiastic	9.241	200.153	221.156	56.290	127.297–454.244
Calm	8.149	184.794	189.347	41.704	85.323–497.137
<i>"Maintenance is vital in preventing many issues and prolonging wood's service life."</i>					
Enthusiastic	4.680	200.642	210.830	50.715	71.326–484.941
Calm	4.753	179.630	179.327	39.360	81.978–488.452
<i>"This is usually accomplished by applying surface finishes that can function as a protective measure, aesthetical measure, or both."</i>					
Enthusiastic	7.606	210.470	228.199	63.590	78.619–487.413
Calm	8.184	200.227	199.871	38.740	79.405–494.667
<i>"Knowing how it works, its possibilities and potential pitfalls, investors and professionals alike can plan and predict ..."</i>					
Enthusiastic	7.052	254.038	252.932	67.838	83.135–494.612
Calm	7.791	203.027	200.071	30.246	76.708–396.422

“... the performance of timber constructions and design them in a way to last and look beautiful for many years to come.”

Enthusiastic	6.459	214.864	230.793	61.355	77.172–468.794
Calm	6.804	195.167	194.586	35.697	90.224–466.280

6.3 Appendix 3: Correlation matrix (Pearson r) of Study 1 outcome variables

	NE	NC	NF	NB	NP	NA	APF	APC	APH	APE	PA	NA	VA	ACT	VAL	INT	MOT	ATT	DIF	EFF	ENJ	LES	ICL	ECL	GCL	ME	K	R	T	SE	KD	RD	TD
NC	-																																
NF	0.07	-																															
NB	0.22	0.33	-																														
NP	0.10	0.07	0.07	-																													
NA	0.04	0.44	0.31	0.04	-																												
APF	0.45	0.09	0.01	0.41	0.54	-																											
APC	0.26	0.20	0.33	0.41	0.44	0.63	-																										
APH	0.42	0.03	0.13	0.34	0.27	0.46	0.47	0.40	-																								
APE	0.49	0.06	0.17	0.41	0.37	0.54	0.63	0.40	0.72	-																							
PA	0.21	0.08	0.05	0.31	0.46	0.37	0.46	0.22	0.12	0.34	-																						
NA	0.04	0.23	0.13	0.21	0.43	0.27	0.38	0.33	0.07	0.23	0.61	-																					
VA	0.13	0.23	0.11	0.28	0.41	0.32	0.43	0.30	0.22	0.29	0.66	0.57	0.53	-																			
ACT	0.12	0.07	0.11	0.29	0.52	0.29	0.44	0.35	0.22	0.29	0.55	0.43	0.66	0.68	-																		
VAL	0.18	0.12	0.04	0.31	0.57	0.39	0.56	0.31	0.16	0.42	0.59	0.13	0.43	0.47	0.46	-																	
INT	0.24	0.01	0.00	0.25	0.60	0.46	0.67	0.45	0.35	0.59	0.56	0.14	0.43	0.53	0.48	0.81	-																
MOT	0.33	0.04	0.03	0.37	0.53	0.40	0.63	0.37	0.23	0.44	0.60	0.28	0.55	0.50	0.51	0.67	0.68	-															
ATT	0.29	0.10	0.01	0.28	0.53	0.40	0.63	0.37	0.23	0.44	0.60	0.28	0.55	0.50	0.51	0.67	0.68	0.43	0.81	-													
DIF	0.01	0.11	0.08	0.08	0.18	0.09	0.11	0.07	0.16	0.02	0.17	0.11	0.11	0.04	-																		
EFF	0.04	0.07	0.06	0.13	0.19	0.20	0.21	0.15	0.11	0.11	0.15	0.20	0.35	0.23	0.10	0.20	0.21	0.43	0.24	0.21	0.20	0.21	0.43	0.43	0.03	-							
ENJ	0.26	0.00	0.07	0.31	0.59	0.38	0.55	0.32	0.13	0.38	0.60	0.42	0.47	0.45	0.81	0.77	0.68	0.24	0.21	0.20	0.21	0.43	0.43	0.03	0.03	-							
LES	0.32	0.09	0.08	0.33	0.53	0.48	0.64	0.43	0.31	0.50	0.51	0.16	0.37	0.43	0.40	0.89	0.77	0.64	0.24	0.21	0.20	0.21	0.43	0.43	0.03	0.03	0.03	-					
ICL	0.13	0.10	0.00	0.08	0.01	0.07	0.14	0.18	0.26	0.17	0.06	0.00	0.06	0.21	0.09	0.01	0.17	0.10	0.44	0.48	0.01	0.10	0.70	0.31	0.16	0.10	0.31	-					
ECL	0.17	0.03	0.20	0.31	0.40	0.27	0.42	0.39	0.16	0.24	0.22	0.14	0.16	0.11	0.25	0.36	0.33	0.32	0.44	0.16	0.35	0.37	0.32	0.33	0.35	0.31	0.31	0.31	-				
GCL	0.11	0.16	0.05	0.12	0.39	0.25	0.42	0.38	0.19	0.26	0.25	0.13	0.28	0.30	0.22	0.41	0.44	0.43	0.01	0.33	0.38	0.37	0.32	0.33	0.35	0.31	0.31	0.31	0.31	-			
ME	0.06	0.04	0.01	0.18	0.29	0.23	0.28	0.20	0.15	0.18	0.33	0.13	0.38	0.48	0.31	0.31	0.36	0.35	0.14	0.59	0.27	0.30	0.36	0.32	0.33	0.35	0.31	0.31	0.31	0.31	-		
K	0.00	0.04	0.12	0.04	0.22	0.04	0.12	0.02	0.04	0.07	0.13	0.17	0.17	0.05	0.07	0.30	0.19	0.24	0.32	0.35	0.37	0.32	0.33	0.35	0.31	0.31	0.31	0.31	0.31	-			
R	0.03	0.04	0.11	0.00	0.26	0.04	0.14	0.02	0.15	0.01	0.06	0.16	0.16	0.02	0.06	0.28	0.21	0.25	0.28	0.32	0.35	0.37	0.32	0.33	0.35	0.31	0.31	0.31	0.31	0.31	-		
T	0.04	0.03	0.10	0.09	0.10	0.08	0.05	0.01	0.15	0.01	0.06	0.16	0.14	0.09	0.06	0.24	0.11	0.14	0.28	0.32	0.35	0.37	0.32	0.33	0.35	0.31	0.31	0.31	0.31	0.31	-		
SE	0.00	0.06	0.02	0.12	0.30	0.26	0.26	0.10	0.13	0.32	0.32	0.19	0.33	0.25	0.40	0.28	0.34	0.28	0.32	0.35	0.37	0.32	0.33	0.35	0.31	0.31	0.31	0.31	0.31	0.31	-		
KD	0.04	0.05	0.07	0.10	0.23	0.12	0.13	0.02	0.04	0.14	0.25	0.03	0.24	0.10	0.21	0.27	0.25	0.22	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	
RD	0.05	0.09	0.07	0.11	0.21	0.06	0.01	0.06	0.16	0.02	0.25	0.03	0.20	0.07	0.27	0.25	0.22	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	
TD	0.00	0.03	0.05	0.06	0.22	0.06	0.01	0.06	0.16	0.02	0.25	0.03	0.20	0.07	0.27	0.25	0.22	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	
SED	0.07	0.10	0.20	0.10	0.22	0.16	0.13	0.09	0.16	0.18	0.33	0.05	0.25	0.28	0.31	0.26	0.27	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	

Note. **$p < 0.001$** ; NE – narrator enthusiasm, NC – narrator calmness, NF – narrator frustration, NB – narrator boredom, NP – narrator pleasantness, NA – narrator activation level, APF – facilitating learning, APC – credible, APH – human-like, APE – engaging, PA – positive activation change score, NA – negative activation change score, VA – valence (PANAVA-KS) change score, ACT – activation level average change score, VAL – valence average change score, INT – situational interest, MOT – intrinsic motivation, ATT – paying attention, DIF – difficulty, EFF – effort, ENJ – enjoyment, LES – more lessons, ICL – intrinsic cognitive load, ECL – extraneous cognitive load, GCL – germane cognitive load, ME – general mental effort, K – knowledge, R – retention, T – transfer, SE – self-evaluation, KD – knowledge (delayed), RD – retention (delayed), TD – transfer (delayed), SED – self-evaluation (delayed)

6.4 Appendix 4: Songs used in Pre-study 2

Song	Title	Author	Duration	Tempo (bpm)	Key
Lively songs					
A	Positive Way	RomanSenykMusic	2:00	111	C major
B	Corporate (Long)	Daddy s Music	4:48	122	C major
F	Fun & Happy	Alex MakeMusic	2:08	125	B$\flat$ major
J	New Happy Corporate (Long)	Daddy s Music	6:00	164	F# major
K	Super Excitement Energy	Musictown	2:23	110	B major
M	Motivational Inspiring Piano	Daddy s Music	1:46	115	A major
N	Beaming With Happiness	lemonmusicstudio	2:09	140	D major
P	Upbeat corporate music for business long	Daddy s Music	4:50	164	F# major
R	Weekend Fun	Muzaproductio	3:02	164	D major
S	Energetic Upbeat Pop Uplifting Corporate (Long)	Daddy s Music	3:55	128	C major
Calm songs					
C	Beautiful Style Ambient	Coma-Media	2:43	100	D major
D	Morning Garden - Acoustic Chill	Olexy	3:53	94	D major
E	Uplifting Ambient	Lesfm	3:16	78	C major
G	Emotional Ambient Cinematic	SoundGalleryBy	2:52	110	C major

	Documentary - Healing Meditation Yoga				
H	Piano Moment	Daddy s Music	4:33	116	D major
I	Warm Corporate	Coma-Media	2:30	110	D major
L	Days To Remember	EvgenyBardyzha	2:18	95	F major
O	Hopeful Slow Atmospheric Meditation	Ashot-Danielyan- Composer	3:55	90	E major
T	Morning Light - Ambient Acoustic Guitar Background Music For Videos	Lesfm	2:59	136	B $\flat$ major
U	Simple Ambient Motivation	Coma-Media	2:28	110	G major

Note. bpm – beats per minute. The ones in bold were selected to be used in Study 2

6.5 Appendix 5: Pre-test questions with answers (correct in bold) in Slovene, English, and Norwegian

PT1 - V Evropi največ škode na lesu povzročijo: / *In Europe, most damage to wood is caused by:* / I Europa er de fleste skadene på skog eller trematerialer forårsaket av:

- a. Hrošči in mravlje / *Beetles and ants* / Biller og maur
- b. Mravlje in termiti / *Ants and termites* / Maur og termitter
- c. Termiti in hrošči / *Termites and beetles* / Termitter og biller**
- d. Mravlje in šurki / *Ants and cockroaches* / Maur og kakerlakker
- e. Ne vem / *I don't know* / Jeg vet ikke

PT2 - Katera vrsta NE spada med mehke lesove? / *PT2 - Which species is NOT a softwood?* / Hvilken art er IKKE et bartre?

- a. Smreka / *Spruce* / Gran
- b. Breza / *Birch* / Bjørk**
- c. Macesen / *Larch* / Lerk
- d. Bor / *Pine* / Furu
- e. Ne vem / *I don't know* / Jeg vet ikke

PT3 - Glulam je: / *Glulam is:* / Limtre er:

- a. Vrsta lesa / *A type of wood* / En tresort
- b. Notranji del debla / *The inner part of the trunk* / Den indre delen av stammen
- c. Lesni kompozit / *A wood composite* / En trekompositt**
- d. Ostanek pri industrijski obdelavi lesa / *Residue from industrial wood processing* / Rester fra treindustrien
- e. Ne vem / *I don't know* / Jeg vet ikke

PT4 - Kaj je poglavitna biološka funkcija lignina? / *What is the main biological function of lignin?* / Hva er den viktigste biologiske funksjonen til lignin?

- a. Zaradi lignina je les trd / *Lignin makes wood hard* / Lignin gjør tre hardt
- b. Zaradi lignina je les rjave barve / *Lignin makes wood brown* / Lignin gjør treet brunt**
- c. Zaradi lignina je les prožen / *Lignin makes wood flexible* / Lignin gjør treet fleksibelt
- d. Zaradi lignina ima les svoj značilen vonj / *Lignin gives wood its characteristic smell* / Lignin gir treet sin karakteristiske lukt
- e. Ne vem / *I don't know* / Jeg vet ikke

PT5 - Les je: / *Wood is*: / Tre:

- a. Dober prevodnik toplote, a slab prevodnik zvoka / *A good conductor of heat but poor conductor of sound* / Leder varme godt, men leder lyd dårlig
- b. Dober prevodnik toplote in zvoka / *A good conductor of heat and sound* / Leder både varme og lyd godt
- c. Slab prevodnik toplote in zvoka / *A poor conductor of heat and sound* / Leder varme og lyd dårlig
- d. **Slab prevodnik toplote, a dober prevodnik zvoka / *A poor conductor of heat but good conductor of sound* / Leder varme dårlig, men leder av lyd godt**
- e. Ne vem / *I don't know* / Jeg vet ikke

PT6 - Za beljavo lesa velja, da je: / *Sapwood is considered to be*: / Splintved/yteved anses å være:

- a. **Manj odporen del lesnega debla / *The less resistant part of the wood trunk* / Den minst motstandsdyktige delen av vedstammen**
- b. Bolj odporen del lesnega debla / *A more resistant part of the wood trunk* / En mest motstandsdyktig delen av vedstammen
- c. Enako odporna kot ostali deli debla / *Equally resistant as other parts of the trunk* / Like motstandsdyktig som andre deler av stammen
- d. Enako odporna kot ostali deli debla, a le pri iglavcih / *Equally resistant as other parts of the trunk, but only in conifers* / Like motstandsdyktig som andre deler av stammen, men bare i bartrær
- e. Ne vem / *I don't know* / Jeg vet ikke

PT7 - Katera izmed naslednjih trditev NE drži? Modifikacija lesa: / *Which of the following statements is NOT true? Modification of wood*: / Hvilket av følgende utsagn er IKKE sant? Modifisering av tre:

- a. Predstavlja okolju prijazen način zaščite lesa / *Is an environmentally friendly way of protecting wood* / Er en miljøvennlig måte å beskytte tre på
- b. Spreminja osnovne lastnosti lesa na molekularnem nivoju / *Modifies the basic properties of wood at the molecular level* / Modifiserer de grunnleggende egenskapene til trevirke på molekylært nivå
- c. **Pomeni nanos vsaj dveh plasti kemične zaščite / *It involves the application of at least two layers of chemical protection* / Innebærer påføring av minst to lag med kjemisk beskyttelse**

- d. Lahko vodi do izgube mase lesa / *Can lead to loss of wood mass* / Kan føre til tap av vedmasse
- e. Ne vem / *I don't know* / Jeg vet ikke

PT8 - Katera izmed naslednjih trditev je pravilna? / *Which of the following statements is correct?* / Hvilket av følgende utsagn er riktig?

- a. Lesene gradnje so zelo nevarne v primeru potresa, saj se les hitro zlomi / *Wooden buildings are very dangerous in the event of an earthquake, as wood breaks quickly* / Trebygninger er svært farlige i tilfelle jordskjelv siden treverk raskt knekker
- b. Lesena gradnja je eden glavnih dejavnikov krčenja gozdov oziroma deforestacije / *Timber construction is one of the main drivers of deforestation* / Trekonstruksjoner er en av hoveddriverne for avskoging
- c. Lesene gradnje so zelo nevarne v primeru požara, saj les zelo hitro zgori / *Wooden buildings are very dangerous in case of fire, as wood burns very quickly* / Trebygninger er svært farlige i tilfelle brann, da ved brenner veldig raskt
- d. **Nobena trditev ni pravilna / *None of the statements is correct* / Ingen av påstandene er riktige**
- e. Ne vem / *I don't know* / Jeg vet ikke

6.6 Appendix 6: Retention and transfer post-test questions with answers (correct in bold) in Slovene, English, and Norwegian

R1 - Lesni hrošči povzročajo največ škode v naslednjem stadiju: / *Wood-boring beetles cause the most damage in their:* / Treborende biller forårsaker mest skade på:

- a. Odrasli / *Adult stage* / Voksenstadiet
- b. Buba / *Pupa stage* / Puppestadiet
- c. **Ličinka / *Larvae stage* / Larvestadiet**
- d. Jajčeca / *Egg stage* / Eggstadiet

R2 - Pri katerem procesu sodelujejo glive? / *In what process are fungi involved?* / I hvilken prosess er sopp involvert?

- a. Preperevanje / *Weathering* / Værpåkjenning
- b. **Trohnenje lesa / *Wood rotting* / Treråte**
- c. Modifikacija lesa / *Wood modification* / Tremodifikasjon
- d. Napadi žuželk / *Insect attacks* / Insektangrep

R3 - Naštejte premaze od najboljšega do najslabšega po odpornosti proti vremenskim vplivom. / *List the coatings from best to worst weathering resistance.* / Hvilken rekkefølge av overflatebehandlinger går fra best til dårligst værbestandighet.

- a. **Barva, lak, olje / *Paint, varnish, oil* / Maling, lakk, olje**
- b. Olje, lak, barva / *Oil, varnish, paint* / Olje, lakk, maling
- c. Lak, olje, barva / *Varnish, oil, paint* / Lakk, olje, maling
- d. Barve, olje, lak / *Paint, oil, varnish* / Maling, olje, lakk

R4 - Razredi odpornosti se nanašajo na / *Durability classes refer to* / Holdbarhetsklasser referer til

- a. **naravno odpornost lesa proti razkroju z glivami, hroščem, termitom in morskim organizmom, ki je opredeljena v evropskem standardu EN 350 (2016) / *the wood's natural resistance to fungal decay, beetles, termites, and marine organisms that are defined in the European standard EN 350 (2016)* / treets naturlige motstand mot råte, biller, termitter og marine organismer som er definert i den europeiske standarden EN 350 (2016)**
- b. največjo odpornost lesa proti razkroju z glivami, hroščem, termitom in morskim organizmom, ki jo je mogoče doseči s postopki modifikacije lesa; opredeljena je v evropskem standardu EN 350 (2016) / *the wood's maximum resistance to fungal decay, beetles, termites, and marine organisms that can be achieved with*

wood modification processes that are defined in the European standard EN 350 (2016) / treets maksimale motstand mot råte, biller, termitter og marine organismer som kan oppnås med tremodifikasjonsprosesser som er definert i den europeiske standarden EN 350 (2016)

- c. *naravno odpornost lesa proti razkroju z glivami, ki je opredeljena v evropskem standardu EN 335 (2013) / the wood's natural resistance to fungal decay, beetles, termites, and marine organisms that are defined in the European standard EN 335 (2013) / treets naturlige motstand mot råte, biller, termitter og marine organismer som er definert i den europeiske standarden EN 335 (2013)*
- d. *najvišje razrede uporabe, ki so priporočeni za določeno vrsto lesa; opredeljeni so v evropskem standardu EN 315 (2013) / the highest usability classes recommended for a particular type of wood; they are defined in the European standard EN 315 (2013) / de høyeste bruksklassene anbefalt for en bestemt tresort, de er definert i den europeiske standarden EN 315 (2013)*

R5 - Kateri od naštetih ukrepov najbolj učinkovito ščiti les v leseni konstrukciji? / *Which of the following is the best protective measure for wood in a wooden structure?* / Hvilket av følgende alternativene er det beste beskyttelsestiltaket for trevirke i en trekonstruksjon?

- a. *Zasnova projekta na način, ki izkorišča izolacijske in akustične lastnosti lesa / Design the project in a way that takes advantage of the insulating and acoustic properties of the wood / Design prosjektet på en måte som utnytter treets isolerende og akustiske egenskaper*
- b. **Zasnova projekta na način, ki omejuje čas, v katerem voda ostane v stiku z lesom / *Design the project in a way that limits the amount of time that water stays in contact with the wood* / Design prosjektet på en måte som begrenser hvor lang tid vannet forblir i kontakt med treverket**
- c. *Uporaba premaza, zaradi katerega les odbija vodo / Use of a coating that makes the wood repel both water / Bruk av et belegg som gjør at treverket avviser vann*
- d. *Izbira lesa, ki je bil podvržen postopku modifikacije, zaradi katerega je les bolj dimenzijsko stabilen / Choosing wood that has undergone a modification process that made the wood more resistant and dimensionally stable / Å velge tre som har gjennomgått en modifikasjonsprosess som har gjort treverket mer motstandsdyktig og formstabilt*

R6 - Življenjska doba stavbe je obdobje med / *A building's service life is the period between* / Et byggs levetid er perioden mellom

- a. gradnjo in rušenjem / *construction and demolition* / bygging og rivning
- b. zasnovo in funkcionalno mejo / *design and functional limit* / design og funksjonell levetid
- c. **gradnjo in funkcionalno mejo / *construction and functional limit* / konstruksjon og funksjonell levetid**
- d. gradnjo in varnostno mejo / *construction and safety limit* / konstruksjon og levetid mht. Sikkerhet

R7 - Večina organizmov, ki uničujejo les, ima naslednje osnovne potrebe: / *Most wood-destroying organisms have the following basic requirements:* / De fleste vedødeleggende organismer har følgende grunnleggende krav:

- a. dovolj vlage, visoke temperature, kisik in hrana / *sufficient moisture, hot temperatures, oxygen, and food* / tilstrekkelig fuktighet, høy temperatur, oksygen og mat
- b. dovolj vlage, visoke temperature in hrana / *sufficient moisture, hot temperatures, and food* / tilstrekkelig fuktighet, høy temperatur og mat
- c. **dovolj vlage, blage temperature, kisik in hrana / *sufficient moisture, mild temperatures, oxygen, and food* / tilstrekkelig fuktighet, mild temperatur, oksygen og mat**
- d. dovolj vlage, blage temperature in kisik / *sufficient moisture, mild temperatures, and oxygen* / tilstrekkelig fuktighet, mild temperatur og oksygen

R8 - Razred uporabe 3 vključuje vse načine uporabe lesa, ki so: / *Use class 3 includes all applications that are:* / Bruksklasse 3 inkluderer alle applikasjoner som er:

- a. Na prostem, v stiku z zemljo ali sladko vodo / *Outside, in direct contact with soil or fresh water* / Utenfor, i direkte kontakt med jord eller ferskvann
- b. Znotraj ali pod streho z občasno nevarnostjo, da se zmoči / *Inside or under cover with an occasional risk of getting wet* / Inne eller under tak og utsatt for fukt
- c. **Na prostem, nad tlemi in izpostavljeni vremenskim vplivom / *Outside, above ground and exposed to weather* / Ute, over bakken og utsatt for vær**
- d. Znotraj in brez nevarnosti, da se zmoči / *Inside and under no risk of getting wet* / Inne og uten risiko for fukt

R9 - Trije glavni procesi propadanja lesa so: / *The three main degradation processes are:* / De tre viktigste nedbrytningsprosessene er:

- a. Poškodbe zaradi vsebnosti soli v zraku, preperevanje, napadi žuželk / *Damage due to salt in the air, weathering, insect infestation* / Skader på grunn av salt i

luften, værpåkjønning, insektangrep

- b. Trohnenje, preperevanje, poškodbe zaradi vlage / *Decay, weathering, moisture damage* / Råte, væråpkjønning, fuktskader
- c. Napadi žuželk, trohnenje, poškodbe zaradi vlage / *Insect infestation, decay, moisture damage* / Insekt angrep, råte, fuktskader
- d. **Trohnenje, preperevanje, napadi žuželk / *Decay, weathering, insect infestation* / Råte, værpåkjønning, insektangrep**

R10 - Prepereli les / Weathered wood / Tre som har vært utsatt for værpåkjønning:

- a. **spremeni barvo, postane bolj grob in izgubi sijaj, nastajati začnejo razpoke / *changes in color, becomes rougher and loses its glossiness, cracks start to form* / endrer farge, får en ru overflate og mister glansen, det begynner å danne seg sprekker**
- b. spremeni barvo, postane žilav in gobast ter znatno izgubi moč / *changes in color, looks stringy and spongy, and loses its strength significantly* / endrer farge, ser trevlete og svampete ut og mister betydelige styrkeegenskaper
- c. je videti beljen, žilav in gobast, nastajati začnejo razpoke / *looks bleached, stringy and spongy, cracks start to form* / ser bleket, trevlet og svampete ut, det begynner å danne seg sprekker
- d. postane temnejši, se skrči, ima kockast videz in znatno izgubi svojo moč / *turns darker, shrinks, breaks into cubes, and loses its strength significantly* / blir mørkere, krymper, brytes i terninger og mister betydelige styrkeegenskaper

R11 - V primerjavi z običajnimi betonskimi in jeklenimi konstrukcijami so zgradbe iz plošč iz križnega lepljenega lesa / *Compared to regular concrete and steel constructions, buildings made of cross laminated timber panels are* / Sammenlignet med vanlige betong- og stålkonstruksjoner er bygninger laget av krysslaminert tre (massivtre)

- a. cenejše, bolj trajnostne, bolj odporne proti ognju in se sestavijo hitreje / *less expensive, more sustainable, more fire resistant, and are assembled faster* / rimeligere, mer bærekraftig, mer brannbestandig og har raskere montasjetid
- b. cenejše, bolj trajnostne in se sestavijo hitreje / *less expensive, more sustainable, and are assembled faster* / rimeligere, mer bærekraftig og har raskere montasjetid
- c. **bolj trajnostne, se sestavijo hitreje in omogočajo višjo stopnjo pred-montaže / *more sustainable, assembled faster, and allow for a higher degree of prefabrication* / mer bærekraftig, har raskere montasjetid, og gir mulighet for en høyere grad av prefabrikasjon**

- d. bolj trajnostne, lahko dostopne in varnejše / *more sustainable, easily available, and safer* / mer bærekraftig, lett tilgjengelig og tryggere

R12 - Stavbe ocenjujemo po več merilih zmogljivosti, ta merila pa imajo svoje meje. Navedite vrstni red teh meja od tiste, ki jo stavba doseže najprej do najkasneje. / *Buildings are evaluated based on several performance criteria, and these have their limits. List the order of these limits from the one the building reaches first to last.* / Bygninger vurderes ut fra flere ytelseskriterier, og disse har sine begrensninger. Hvilken liste har riktig rekkefølgen på disse grensene, fra grensen som bygningen når først til den grensen som bygningen når sist.

- a. Funkcionalna meja, varnostna meja, estetska meja / *Functional limit, safety limit, aesthetic limit* / Funksjonsgrense, sikkerhetsgrense, estetisk grense
- b. Varnostna meja, funkcionalna meja, estetska meja / *Safety limit, functional limit, aesthetic limit* / Sikkerhetsgrense, funksjonsgrense, estetisk grense
- c. Varnostna meja, estetska meja, funkcionalna meja / *Safety limit, aesthetic limit, functional limit* / Sikkerhetsgrense, estetisk grense, funksjonsgrense
- d. **Estetska meja, funkcionalna meja, varnostna meja / *Aesthetic limit, functional limit, safety limit* / Estetisk grense, funksjonsgrense, sikkerhetsgrense**

R13 - Glavni vzrok trohnenja lesa je / *The leading cause of wood rot is* / Den viktigste årsaken til råte er

- a. kombinacija gliv in visokih temperatur / *fungi and high temperatures* / sopp og høye temperaturer
- b. dlje časa trajajoča vlaga / *moisture for longer periods of time* / fuktighet over lengre tid
- c. **kombinacija gliv in dlje časa trajajoče vlage / *fungi and moisture for longer periods of time* / sopp og fukt i lengre perioder**
- d. visoke temperature / *high temperatures* / høye temperature

R14 - Lesne hrošče še posebej privlačijo / *Wood-boring beetles are especially attracted to* / Treborende biller tiltrekkes av

- a. vrste lesa z nižjo vsebnostjo škroba / *wood with lower starch content* / tre med lavere stivelsesinnhold
- b. **grobe površine z razpokami / *rough surfaces with cracks* / ru overflater med sprekker**
- c. hladna in suha območja / *cold and dry areas* / kalde og tørre områder

- d. zelo vlažna in svetla območja / *highly humid and light areas* / svært fuktige og lyse områder

R15 - Smreko in hrast uvrščamo v naslednje razrede odpornosti proti razkroju z glivami:
// *In terms of fungal decay, spruce and oak are classified as:* / Når det gjelder råte, er gran og eik klassifisert som:

- a. **Hrast je odporen (RU2), smreka je neodporna (RU4) / Oak is durable (DC2), spruce is slightly durable (DC4) / Eik er holdbar (DC2), gran er litt holdbar (DC4)**
- b. Obe vrsti spadeta med odporne vrste (RU2) / *Both species are durable (DC2)* / Begge artene er holdbare (DC2)
- c. Obe vrsti spadeta med neodporne vrste (RU4) / *Both species are slightly durable (DC4)* / Begge artene er litt holdbare (DC4)
- d. Smreka je odporna (RU2), hrast je neodporen (RU4) / *Spruce is durable (DC2), oak is slightly durable (DC4)* / Gran er holdbar (DC2), eik er litt holdbar (DC4)

R16 - Lesni hrošči so prisotni / *Wood-boring beetles are present* / Treborende biller finnes

- a. v Sredozemlju / *in the Mediterranean* / ved Middelhavet
- b. po vsej Evropi / all over Europe / over hele Europa**
- c. v srednji Evropi / *in central Europe* / i Sentral-Europa
- d. v skandinavskih državah / *in Scandinavian countries* / i skandinaviske land

R17 - Trohnenje lesa je najbolj izrazito v / *Fungal decay is the most prominent in* / Råte er mest fremtredende i

- a. toplih in suhih regijah / *warm and dry regions* / varme og tørre områder
- b. zmernih območjih z močnimi padavinami / temperate regions with heavy rainfall / tempererte områder med mye nedbør**
- c. Sredozemlju / *Mediterranean* / Middelhavsområdet
- d. višje ležečih območjih z veliko snega / *elevated areas with lots of snow* / I høyfjellet der det er mye snø

R18 - Razredi odpornosti proti razkroju z glivami se določijo s testiranjem kolov, narejenih iz / *Durability classes for decay are determined by testing stakes made from* / Holdbarhetsklasser for nedbryting av tre bestemmes ved å teste staker laget av

- a. **jedrovine, notranjega in mrtvega dela lesa / heartwood, the inner and dead**

part of the wood / kjerneved, den indre og døde delen av veden

- b. jedrovine, zunanjega in mlajšega dela lesa / *heartwood, the outer and younger part of the wood* / kjerneved, den ytre og yngre delen av veden
- c. beljave, zunanjega in mlajšega dela lesa / *sapwood, the outer and younger part of the wood* / yteved, den ytre og yngre delen av treverket
- d. beljave, notranjega in mlajšega dela lesa / *sapwood, the inner and younger part of the wood* / yteved, den indre og yngre delen av treverket

R19 - Preperevanje vpliva na / *Weathering affects* / Værpåkjønning påvirker

- a. varnost stavbe / *the safety of the building* / sikkerheten til bygningen
- b. funkcionalnost stavbe / *the functionality of the building* / funksjonaliteten til bygget
- c. **estetiko stavbe / *aesthetics of the building* / bygningens estetikk**
- d. nič od naštetega / *nothing of the above* / ingenting av det ovennevnte



T1 - Les na fotografiji kaže znake / *The wood on the photo shows signs of* / Treverket på bildet viser tegn til

- a. razkrajjanja zaradi gliv rjave trohnobe / *brown-rot fungi decay* / nedbryting på grunn av brunråtesopp
- b. napada termitov / *termite attack* / termittangrep
- c. **napada lesnih hroščev / *wood-boring beetle infestation* / vedborende billeangrep**
- d. razkrajjanja zaradi gliv bele trohnobe / *white-rot fungi decay* / nedbryting på grunn av hvitråtesopp



T2 - Ta fotografija lesene konstrukcije je bila posneta v obmorskem mestu na Portugalskem. Kaj je glavni razlog za škodo? / *This photo of a wooden structure was taken in a seaside town in Portugal. What is the main reason for the damage?* / Dette bildet av en trekonstruksjon ble tatt

i en kystby i Portugal. Hva er hovedårsaken til skaden?

- a. Napad termitov / *Termite infestation* / Termittangrep
- b. Degradacija zaradi visoke izpostavljenosti soncu / *Degradation due to the amount of sun exposure* / Nedbrytning på grunn av mengden soleksponering
- c. **Trohnenje lesa zaradi zadrževanja vode / *Wood rot due to water staying trapped* / Råte på grunn av at vann samles opp i veden**
- d. Degradacija zaradi vsebnosti soli v zraku / *Degradation due to salt in the air* / Nedbrytning på grunn av salt i luften



T3 - Lastnik se je odločil, da je čas za poživitev fasade. Kaj bi bilo v tem konkretnem primeru najbolj smiselno narediti? / *The owner decided it is time to revitalize the facade. What would be the most sensible way to do so in this specific case?* / Eieren har bestemt seg for at det er på tide å fornye fasaden. Hva vil være den mest fornuftige måten å gjøre det på i dette spesifikke tilfellet?

- a. Očistiti površino, popolnoma pobrusiti obstoječo plast premaza, nanesti novo plast premaza / *Clean the surface, completely sand off the existing layer of coating, apply a new coating* / Rengjør overflaten, slip helt av det eksisterende belegglaget, påfør ny overflatebehandling
- b. **Očistiti površino, nanesti novo plast premaza / *Clean the surface, apply a new coating* / Rengjør overflaten, påfør ny overflatebehandling**
- c. Nanesti novo plast premaza / *Apply a new coating* / Påfør ny overflatebehandling
- d. Zamenjati vse lesene elemente na fasadi / *Replace all wooden elements in the facade* / Skift alle treelementer i fasaden



T4 - Na fotografiji lahko vidimo jasne znake / *On the photo we can see clear signs of* / På bildet kan vi se tydelige tegn til

- a. **preperelosti / *weathering* / skade fra værtpåkjenning**
- b. trohnenja lesa / *wood rot* / råte
- c. dejavnosti žuželk / *insect activity* / insektaktivitet
- d. preperelosti, trohnenja lesa in dejavnosti žuželk / *weathering, decay and insect activity* / skade fra værtpåkjenning, forråtnelse og insektaktivitet

T5 - Termiti so bili opaženi v/na / *Termites have been spotted in* / Termitter er oppdaget i

- a. **Španiji / *Spain* / Spania**
- b. Norveškem / *Norway* / Norge
- c. Španiji in na Norveškem / *Spain and Norway* / Spania og Norge
- d. severovzhodnem delu Slovenije / *northeastern part of Slovenia* / Nordøstlige del av Slovenia

T6 - Kaj bi se zgodilo, če bi več stavb izdelovali iz lesa? / *What would happen if more buildings were made of wood?* / Hva ville skje hvis flere bygninger ble laget av tre?

- a. Skrčili bi ogromno gozdov, kar bi vodilo do porušenja številnih ekosistemov / *Many forests would be cut down, which would lead to the destruction of many ecosystems* / Mange skoger ville bli hogd ned, noe som ville føre til ødeleggelse av mange økosystemer
- b. **Gradbeni sektor bi zmanjšal število emisij in porabljenih virov / *The construction sector would reduce emissions and resources used* / Byggesektoren vil redusere utslipp og ressursbruk**
- c. Ustavili bi klimatske spremembe / *We would stop climate change* / Vi ville stoppe klimaendringene

- d. Stavbe, še posebej tiste z večjim številom nadstropij, bi postale manj varne za uporabnike / *Buildings, especially multi-storey ones, would become less safe dangerous for users* / Bygninger, spesielt fleretasjes bygg, vil bli mindre trygge og farlige for brukerne

T7 - Moderna gradnja iz lesa večinoma uporablja / *Modern wood construction mostly uses* / Moderne trekonstruksjon bruker for det meste

- a. les iglavcev / *wood from coniferous tree species* / tre fra bartreslag
b. les listavcev / *wood from deciduous tree species* / ved fra løvtreslag
c. odporen les iz tropskih gozdov / *durable wood from tropical forests* / holdbart treverk fra tropiske skoger
d. **inženirske lesne izdelke / *engineered wood products* / sammensatte (engineered) treprodukter**



T8 - Kje se bo trohnoba najverjetneje pojavila najprej? / *Where will wood rot most likely occur first?* / Hvor vil råde mest sannsynlig oppstå først?

- a. Na južni strani, saj tam sije sonce najdlje / *On the south side, where the sun shines the longest* / På sørsiden, der solen skinner lengst
b. Na stiku med strešno kritino in lesom / *At the junction between the roofing and wood* / I overgangen mellom takteking og tre
c. **Na stiku z zemljo, ker koča ni dvignjena od tal / *In contact with the soil, as the hut is not raised of the ground* / I jordkontakt, siden hytta ikke er hevet av bakken**
d. Med lesenimi deskami, saj med njimi ni dovolj prostora / *Between the wooden boards, as there is not enough space between them* / Mellom kledningsbordene, fordi det ikke er nok plass mellom dem



T9 - Čeprav sta iz iste vrste lesa, je rdeči del lesenega tramu v slabšem stanju kot moder. Zakaj? / *Although they are made of the same type of wood, the red part of the wooden beam is in worse shape than the blue one. Why?* / Selv om de er laget av samme tresort, er den røde delen av trebjelken i dårligere forfatning enn den blå. Hvorfor?

- a. Ker je bolj izpostavljen soncu / *Because it is more exposed to the sun* / Fordi den er mer utsatt for solen
- b. Ker čelni del lesenega tramu vpije več vlage / *Because end-grains absorb more moisture* / Fordi endaved absorberer mer fuktighet**
- c. Ker je bolj izpostavljen dežju / *Because it is more exposed to rain* / Fordi det er mer utsatt for regn
- d. Ker je postavljen vodoravno / *Because it is placed horizontally* / Fordi den er plassert horisontalt

T10 - Preperevanje je mogoče popolnoma ustaviti / *Weathering can be completely stopped with* / Effekten fra værpåkjennning kan stoppes helt med

- a. s premazi, ki tvorijo film / *film-forming coatings* / filmdannende belegg
- b. s prodornimi premazi / *penetrating coatings* / gjennomtrengende belegg
- c. z obojimi / *both* / belegg beleggtypene
- d. z ničimer od naštetega / *none of the above* / ingen av de ovennevnte**

6.7 Appendix 7: Normality and homogeneity test results for Study 1 outcomes

Table 123: Shapiro-Wilk's normality and Levene's homogeneity tests for Study 1 outcome variables before comparisons

Item	<i>Enthusiastic vs. calm voice</i>				<i>No SLS vs. SLS</i>			
	<i>F</i>	<i>p</i>	<i>W</i>	<i>p</i>	<i>F</i>	<i>p</i>	<i>W</i>	<i>p</i>
Narrator affective state*								
Enthusiasm	1.408	0.237	0.949	< .001	0.962	0.328	0.898	< .001
Calmness	2.168	0.142	0.916	< .001	0.135	0.714	0.859	< .001
Frustration	0.728	0.394	0.598	< .001	0.774	0.380	0.599	< .001
Boredom	0.786	0.376	0.963	< .001	0.075	0.785	0.938	< .001
Pleasantness	1.497	0.222	0.973	< .001	0.435	0.510	0.952	< .001
Activation I.	0.949	0.331	0.968	< .001	0.045	0.832	0.966	< .001
Narrator perception*								
Facilitating learning	10.885	0.001	0.993	0.414	0.190	0.663	0.952	< .001
Credibility	13.368	< .001	0.969	< .001	0.076	0.783	0.979	0.002
Human-like	0.113	0.737	0.983	0.008	0.072	0.789	0.970	< .001
Engaging	1.163	0.282	0.983	< .001	4.613	0.033	0.987	0.045
Participants' affective state								
	<i>Enthusiastic narrator group</i>				<i>No SLS group</i>			
PA			0.980	0.103			0.973	0.018
NA			0.982	0.149			0.977	0.050
VA			0.974	0.027			0.976	0.036
Activation lev. 1			0.890	< .001			0.943	< .001
Activation lev. 2			0.945	0.001			0.915	< .001
Activation lev. 3			0.962	0.003			0.955	0.003
Activation lev. 4			0.956	0.001			0.961	0.002
Activation lev. 5			0.973	0.022			0.970	0.011
Activation level average			0.977	0.057			0.971	0.014

Valence 1	0.918	< .001	0.879	< .001
Valence 2	0.943	< .001	0.940	< .001
Valence 3	0.939	< .001	0.947	< .001
Valence 4	0.950	< .001	0.961	0.002
Valence 5	0.944	< .001	0.938	< .001
Valence average	0.978	0.069	0.976	0.037
<i>Calm narrator group</i>		<i>SLS group</i>		
PA	0.945	< .001	0.947	< .001
NA	0.960	0.002	0.953	< .001
VA	0.944	< .001	0.952	< .001
Activation I. 1	0.934	< .001	0.884	< .001
Activation I. 2	0.912	< .001	0.938	< .001
Activation I. 3	0.956	< .001	0.961	0.002
Activation I. 4	0.965	0.004	0.965	0.005
Activation I. 5	0.959	0.001	0.968	0.008
Activation level average	0.977	0.042	0.985	0.223
Valence 1	0.928	< .001	0.942	< .001
Valence 2	0.942	< .001	0.941	< .001
Valence 3	0.926	< .001	0.941	< .001
Valence 4	0.957	0.001	0.940	< .001
Valence 5	0.948	< .001	0.953	< .001
Valence average	0.969	0.010	0.970	0.013
Interest and motivation				
Situational interest [†]	0.313	0.576	0.983	0.007
Interest (delayed) [§]	0.801	0.373	0.939	< .001
Intrinsic motivation [†]	4.085	0.044	0.984	0.012
Learners' experience [†]				

Paying attention	4.446	0.036	0.945	< .001	4.030	0.046	0.940	< .001
Difficulty	0.299	0.585	0.963	< .001	0.147	0.701	0.953	< .001
Exerting more effort	0.959	0.329	0.940	< .001	2.263	0.134	0.946	< .001
Enjoyment	4.405	0.037	0.951	< .001	1.344	0.248	0.958	< .001
More lessons	0.424	0.121	950	< .001	1.722	0.191	0.960	< .001
Cognitive outcomes								
Intrinsic cognitive load [†]	0.438	0.509	0.979	0.002	0.999	0.319	0.978	0.002
Extraneous cognitive load [†]	1.754	0.187	0.986	0.025	0.046	0.830	0.982	0.005
Germane cognitive load [†]	0.005	0.942	0.970	< .001	5.420	0.021	0.978	0.001
Mental effort 1*	0.619	0.432	0.974	< .001	0.523	0.470	0.966	< .001
Mental effort 2*	2.294	0.131	0.978	0.001	1.154	0.284	0.980	0.003
Mental effort 3*	0.004	0.948	0.963	< .001	1.022	0.313	0.961	< .001
Mental effort 4*	1.058	0.305	0.943	< .001	0.010	0.921	0.949	< .001
Mental effort 5*	0.519	0.472	0.948	< .001	0.001	0.979	0.965	< .001
Mental effort average*	0.128	0.721	0.979	0.002	0.046	0.830	0.979	0.002
Learning outcomes								
<i>Immediate testing[‡]</i>								
Knowledge	0.610	0.436	0.610	0.436	9.705	0.002	0.971	< .001
Retention	2.133	0.146	2.133	0.146	8.547	0.004	0.976	< .001
Transfer	0.338	0.562	0.338	0.562	1.555	0.214	0.984	0.013
Certainty	7.917	0.005	7.917	0.005	0.022	0.883	0.988	0.059
Certainty in correct answers	6.004	0.015	6.004	0.015	0.022	0.881	0.982	0.007
Certainty in incorrect answers	9.126	0.003	9.126	0.003	0.410	0.523	0.991	0.159
R Certainty	6.463	0.012	6.463	0.012	0.020	0.887	0.989	0.078
R Certainty in correct answers	2.983	0.086	2.983	0.086	0.037	0.848	0.983	0.007

R Certainty in incorrect answers	5.572	0.019	5.572	0.019	0.380	0.538	0.992	0.234
T Certainty	10.791	0.001	10.791	0.001	0.149	0.700	0.984	0.014
T Certainty in correct answers	7.091	0.008	7.091	0.008	0.032	0.859	0.979	0.002
T Certainty in incorrect answers	10.251	0.002	10.251	0.002	0.006	0.940	0.992	0.311
Self-evaluation	1.111	0.293	1.111	0.293	0.690	0.407	0.927	< .001
<i>Delayed testing[§]</i>								
Knowledge	1.041	0.310	0.968	0.021	7.402	0.008	0.970	0.031
Retention	0.455	0.502	0.975	0.067	3.585	0.061	0.977	0.103
Transfer	0.387	0.535	0.981	0.202	12.369	< .001	0.980	0.164
Certainty	1.904	0.171	0.985	0.323	0.009	0.924	0.986	0.420
Certainty in correct answers	1.135	0.289	0.987	0.460	0.379	0.540	0.985	0.342
Certainty in incorrect answers	1.682	0.198	0.985	0.327	0.081	0.777	0.986	0.415
R Certainty	1.730	0.192	0.980	0.139	0.013	0.911	0.985	0.334
R Certainty in correct answers	1.141	0.288	0.989	0.629	0.717	0.399	0.986	0.425
R Certainty in incorrect answers	1.463	0.230	0.977	0.085	1.448	0.232	0.987	0.488
T Certainty	2.158	0.145	0.985	0.321	0.276	0.601	0.987	0.474
T Certainty in correct answers	0.787	0.377	0.982	0.209	0.526	0.470	0.981	0.185
T Certainty in incorrect answers	1.459	0.230	0.985	0.340	1.464	0.229	0.989	0.593
Self-evaluation	0.771	0.382	0.885	< .001	0.013	0.910	0.885	< .001

Note. PA – positive activation, NA – negative activation, VA – valence, R – retention, T – transfer; * $df_1 = 1$, $df_2 = 224$; † $df_1 = 1$, $df_2 = 223$; ‡ $df_1 = 1$, $df_2 = 222$; § $df_1 = 1$, $df_2 = 94$

6.8 Appendix 8: Descriptive statistics for main outcomes by low proficiency narrator group

Table 124: Descriptive statistics for the main outcome variables for enthusiastic and calm narrator groups on the lower English proficiency group (LexTALE < 63)

	Group	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>	Skew- ness	Kurto- sis
Narrator emotional tone							
Enthusiasm	Enthusiastic	3.56	1.69	1.00	7.00	−0.11	−0.87
	Calm	2.26	1.62	1.00	7.00	1.03	0.12
Calmness	Enthusiastic	5.32	1.36	2.00	7.00	−0.23	−0.85
	Calm	6.17	1.20	3.00	7.00	−1.16	−0.12
Frustration	Enthusiastic	1.66	1.14	1.00	6.00	1.80	2.88
	Calm	1.69	1.27	1.00	6.00	1.67	1.58
Boredom	Enthusiastic	3.02	1.77	1.00	7.00	0.54	−0.50
	Calm	4.12	1.96	1.00	7.00	−0.00	−1.05
Pleasantness	Enthusiastic	4.63	1.46	1.00	7.00	−0.66	0.02
	Calm	4.09	1.53	1.00	7.00	0.13	−0.36
Activation level	Enthusiastic	3.76	1.32	1.00	7.00	−0.11	−0.06
	Calm	3.29	1.38	1.00	7.00	0.20	−0.17
Social partnership with the narrator							
Facilitating learning	Enthusiastic	4.21	0.87	2.00	5.90	−0.32	−0.02
	Calm	3.85	1.20	1.20	6.10	−0.28	−0.22
Credibility	Enthusiastic	5.29	0.85	3.00	7.00	−0.32	0.10
	Calm	4.84	1.25	1.00	7.00	−0.69	0.46
Human-like	Enthusiastic	4.44	1.38	1.00	6.40	−0.80	−0.27
	Calm	3.87	1.42	1.00	6.80	−0.06	−0.52
Engaging	Enthusiastic	3.48	1.32	1.00	6.00	0.09	−0.53
	Calm	3.01	1.45	1.00	6.80	0.57	−0.19
Differences in affective state							
PA change score	Enthusiastic	−0.58	1.33	−4.00	2.25	−0.20	0.26
	Calm	−0.51	1.04	−4.50	2.00	−1.03	3.27

NA change score	Enthusiastic	−0.06	0.88	−2.00	2.25	0.12	0.22
	Calm	−0.37	1.08	−3.25	2.00	−0.50	0.40
VA change score	Enthusiastic	−0.41	1.27	−3.00	2.50	0.09	0.11
	Calm	−0.18	1.22	−4.00	3.00	−0.81	2.54
Activation level change score	Enthusiastic	−0.42	1.36	−3.60	3.60	0.06	0.60
	Calm	−0.09	1.45	−3.80	4.00	0.42	1.43
Valence change score	Enthusiastic	−0.65	1.51	−5.60	2.40	−0.87	1.03
	Calm	−0.18	1.65	−4.40	3.40	−0.38	0.29
Interest and motivation							
Situational interest	Enthusiastic	3.31	1.17	1.00	5.83	−0.09	−0.32
	Calm	3.60	1.16	1.00	5.50	−0.42	−0.47
Interest (delayed)	Enthusiastic	3.35	1.32	1.00	6.00	−0.03	−0.73
	Calm	3.72	1.03	2.00	5.00	−0.24	−1.05
Intrinsic motivation	Enthusiastic	3.59	1.19	1.00	5.88	−0.40	−0.34
	Calm	3.52	1.29	1.00	6.13	−0.19	−0.67
Learners' experience							
Paying attention	Enthusiastic	3.46	1.39	1.00	6.00	−0.01	−0.93
	Calm	3.29	1.39	1.00	6.00	0.14	−0.79
Difficulty	Enthusiastic	4.00	1.56	1.00	7.00	0.11	−0.72
	Calm	3.86	1.39	1.00	7.00	−0.11	−0.66
Exerting more effort	Enthusiastic	3.81	1.48	1.00	7.00	−0.13	−0.24
	Calm	3.72	1.44	1.00	6.00	−0.01	−1.10
Enjoyment	Enthusiastic	3.61	1.38	1.00	7.00	0.01	−0.38
	Calm	3.59	1.56	1.00	7.00	−0.25	−0.73
More lessons like this	Enthusiastic	3.36	1.53	1.00	7.00	0.36	−0.31
	Calm	3.38	1.68	1.00	7.00	0.17	−0.61
Cognitive outcomes							
Intrinsic cognitive load	Enthusiastic	4.38	1.11	2.00	6.50	−0.09	−0.60
	Calm	4.02	1.21	1.50	6.00	−0.22	−0.61
Extraneous cognitive load	Enthusiastic	3.66	1.24	1.67	6.67	0.56	−0.49
	Calm	3.85	1.23	1.33	7.00	0.11	−0.49

Germane cognitive load	Enthusiastic	4.46	1.19	2.00	6.50	-0.48	-0.52
	Calm	4.67	1.23	1.00	7.00	-0.82	0.58
Mental effort (average)	Enthusiastic	4.92	1.36	1.40	7.80	-0.30	0.40
	Calm	4.66	1.48	1.80	8.80	0.01	-0.06
Immediate testing							
Knowledge	Enthusiastic	12.68	4.00	5.00	23	0.48	-0.18
	Calm	11.95	3.64	5.00	24.00	0.83	1.65
Retention	Enthusiastic	8.46	2.81	2.00	16.00	0.40	0.32
	Calm	7.88	2.58	3.00	15.00	0.47	0.49
Transfer	Enthusiastic	4.22	1.92	0.00	8.00	-0.11	-0.09
	Calm	4.07	1.95	0.00	9.00	0.21	-0.51
Certainty	Enthusiastic	50.55	16.47	8.62	87.03	-0.21	-0.20
	Calm	47.04	21.73	6.66	88.59	-0.02	-0.84
Certainty in correct answers	Enthusiastic	52.94	17.39	10.00	92.50	-0.09	-0.37
	Calm	51.17	23.64	6.15	93.33	-0.03	-1.00
Certainty in incorrect answers	Enthusiastic	48.29	16.33	7.00	79.08	-0.43	0.08
	Calm	44.08	20.89	0.00	84.61	-0.10	-0.74
R Certainty	Enthusiastic	49.72	17.15	9.21	90.00	-0.01	-0.45
	Calm	47.03	21.51	6.47	88.37	-0.02	-0.94
R Certainty in correct answers	Enthusiastic	52.82	19.21	12.50	96.43	0.11	-0.65
	Calm	50.58	22.75	6.11	93.33	-0.00	-1.00
R Certainty in incorrect answers	Enthusiastic	46.84	16.50	6.82	85.57	-0.11	-0.26
	Calm	44.10	21.21	0.00	85.56	-0.12	-0.84
T Certainty	Enthusiastic	52.12	17.61	7.50	87.00	-0.36	-0.04
	Calm	47.07	24.96	0.00	92.50	0.04	-0.85
T Certainty in correct answers	Enthusiastic	53.11	19.03	6.00	97.50	-0.04	-0.30
	Calm	50.32	28.06	0.00	100.00	0.02	-1.13
T Certainty in incorrect answers	Enthusiastic	51.50	19.13	0.00	100.00	-0.18	0.48
	Calm	44.97	24.14	0.00	100.00	0.19	-0.59
Self-evaluation	Enthusiastic	3.14	0.94	1.00	6.00	-0.15	0.76

	Calm	3.50	1.22	1.00	7.00	−0.03	0.76
Delayed testing							
Knowledge	Enthusiastic	12.58	3.92	4.00	21.00	0.14	0.28
	Calm	10.34	3.55	4.00	18.00	0.19	−0.67
Retention	Enthusiastic	8.15	2.95	2.00	14.00	0.10	−0.03
	Calm	6.86	2.45	3.00	12.00	0.33	−0.73
Transfer	Enthusiastic	4.42	1.84	0.00	7.00	−0.39	−0.25
	Calm	3.48	1.53	1.00	7.00	0.20	−0.38
Certainty	Enthusiastic	39.80	20.88	0.00	77.93	−0.28	−0.41
	Calm	43.44	20.75	0.00	79.31	−0.26	−0.55
Certainty in correct answers	Enthusiastic	41.51	20.93	0.00	78.89	−0.36	−0.31
	Calm	46.74	21.31	0.00	82.56	−0.41	−0.50
Certainty in incorrect answers	Enthusiastic	38.55	21.06	0.00	77.50	−0.17	−0.53
	Calm	41.28	20.87	0.00	80.36	−0.12	−0.50
R Certainty	Enthusiastic	38.21	20.51	0.00	74.74	−0.19	−0.37
	Calm	41.05	20.27	0.00	76.32	−0.16	−0.64
R Certainty in correct answers	Enthusiastic	40.07	21.10	0.00	78.89	−0.13	−0.47
	Calm	43.69	20.96	0.00	83.67	−0.21	−0.53
R Certainty in incorrect answers	Enthusiastic	36.17	20.28	0.00	71.00	−0.12	−0.44
	Calm	39.52	20.55	0.00	77.78	−0.03	−0.66
T Certainty	Enthusiastic	42.82	22.53	0.00	84.00	−0.27	−0.53
	Calm	47.99	22.44	0.00	85.00	−0.36	−0.57
T Certainty in correct answers	Enthusiastic	41.98	22.25	0.00	77.50	−0.42	−0.53
	Calm	51.50	24.14	0.00	85.00	−0.39	−0.75
T Certainty in incorrect answers	Enthusiastic	42.26	23.04	0.00	84.00	−0.05	−0.55
	Calm	44.43	23.07	0.00	85.00	−0.09	−0.81
Self-evaluation	Enthusiastic	2.96	0.87	1.00	4.00	−0.71	0.24
	Calm	3.41	0.98	1.00	5.00	−0.71	−0.10

Note. PA – positive activation, NA – negative activation, VA – valence, R – retention, T – transfer

6.9 Appendix 9: ANCOVA comparisons by low proficiency narrator group

Table 125: ANCOVA comparisons between the enthusiastic and calm narrator on all main dependable variables on the lower English proficiency group (LexTALE < 63)

	ANCOVA*			Homogeneity test**		Normality test	
	<i>F</i>	<i>p</i>	η^2p	<i>F</i>	<i>p</i>	<i>W</i>	<i>p</i>
Narrator emotional tone							
Enthusiasm	18.073	< .001	0.142	0.079	0.779	0.963	0.003
Calmness	12.186	< .001	0.101	0.268	0.606	0.957	< .001
Frustration	0.107	0.744	0.001	1.626	0.205	0.858	< .001
Boredom	11.084	0.001	0.092	1.506	0.222	0.960	0.001
Pleasantness	5.375	0.022	0.047	0.063	0.803	0.986	0.266
Activation level	5.882	0.017	0.051	0.007	0.933	0.988	0.402
Social partnership with the narrator							
Facilitating learning ^Q	3.321	0.071		5.214	0.024	0.991	0.610
Credibility ^Q	4.14	0.044		10.17 3	0.002	0.980	0.077
Human-like	4.304	0.040	0.038	0.416	0.520	0.985	0.202
Engaging	3.917	0.050	0.035	0.528	0.469	0.982	0.123
Differences in affective state							
Positive activation	0.233	0.631	0.002	0.052	0.820	0.986	0.255
Negative activation	1.107	0.295	0.010	0.245	0.621	0.993	0.862
Valence	1.478	0.227	0.013	0.800	0.373	0.979	0.068
Activation level	1.375	0.243	0.012	0.018	0.893	0.982	0.116
Valence	0.475	0.492	0.004	0.037	0.848	0.952	< .001
Interest and motivation							
Situational interest	0.960	0.329	0.009	1.020	0.315	0.985	0.208
Interest (delayed)	0.101	0.752	0.002	0.007	0.933	0.980	0.506
Intrinsic motivation	0.814	0.369	0.007	0.276	0.601	0.986	0.256
Learners' experience							
Paying attention	0.947	0.333	0.009	0.052	0.819	0.987	0.341

Difficulty	0.182	0.671	0.002	2.010	0.159	0.986	0.256
Exerting more effort	0.197	0.658	0.002	0.001	0.972	0.987	0.325
Enjoyment	0.109	0.742	0.001	0.000	0.992	0.994	0.906
More lessons like this	0.040	0.843	0.000	0.010	0.922	0.979	0.059
Cognitive outcomes							
Intrinsic cognitive load	3.551	0.062	0.032	0.264	0.608	0.979	0.062
Extraneous cognitive load	0.579	0.448	0.005	0.162	0.688	0.978	0.049
Germane cognitive load	0.307	0.580	0.003	0.180	0.672	0.970	0.011
Mental effort (average)	1.578	0.212	0.014	0.842	0.361	0.983	0.136
Immediate testing							
Knowledge	0.439	0.509	0.004	1.860	0.175	0.983	0.140
Retention	0.685	0.410	0.006	0.392	0.533	0.987	0.331
Transfer	0.010	0.919	0.000	0.077	0.782	0.992	0.749
Certainty ^Q	0.592	0.443		4.275	0.041	0.987	0.327
Certainty in correct answers ^Q	0.023	0.881		6.616	0.011	0.988	0.394
Certainty in incorrect answers ^Q	1.205	0.275		4.449	0.037	0.988	0.426
R Certainty ^Q	0.306	0.581		3.585	0.061	0.989	0.431
R Certainty in correct answers ^Q	0.191	0.663	0.002	1.403	0.239	0.989	0.455
R Certainty in incorrect answers ^Q	0.317	0.575		5.301	0.023	0.989	0.429
T Certainty ^Q	1.138	0.288		5.136	0.025	0.989	0.431
T Certainty in correct answers ^Q	0.061	0.805		11.388	0.001	0.985	0.223
T Certainty in incorrect answers	2.050	0.155	0.018	1.749	0.189	0.991	0.602
Self-evaluation ^Q	4.554	0.035		4.102	0.045	0.976	0.031
Delayed testing [†]							
Knowledge	8.282	0.006	0.150	1.007	0.320	0.983	0.638
Retention	4.563	0.038	0.088	3.217	0.079	0.986	0.751
Transfer	8.567	0.005	0.154	2.150	0.148	0.987	0.804

Certainty	0.373	0.544	0.008	0.029	0.866	0.982	0.601
Certainty in correct answers	0.809	0.373	0.018	0.019	0.890	0.984	0.675
Certainty in incorrect answers	0.135	0.715	0.003	0.118	0.733	0.984	0.676
R Certainty	0.235	0.630	0.005	0.001	0.974	0.979	0.474
R Certainty in correct answers	0.374	0.544	0.008	0.000	0.990	0.981	0.572
R Certainty in incorrect answers	0.259	0.613	0.006	0.055	0.815	0.974	0.299
T Certainty	0.656	0.422	0.014	0.159	0.692	0.987	0.832
T Certainty in correct answers	2.496	0.121	0.054	0.063	0.803	0.975	0.340
T Certainty in incorrect answers	0.034	0.854	0.001	0.266	0.609	0.988	0.870
Self-evaluation	2.230	0.142	0.045	0.477	0.493	0.978	0.405

Note. * $df_1 = 1$, $df_2 = 109$; ** $df_1 = 1$, $df_2 = 115$; † * $df_1 = 1$, $df_2 = 47$; ** $df_1 = 1$, $df_2 = 53$; R – retention, T – transfer, ^Q – Quade test results reported instead of ANCOVA due to the homogeneity of variances assumption not being met; covariates included are prior tested knowledge, prior interest, LexTALE score, and PANAVA-KS baseline measures, except for the Activation level and Valence variables, which had the activation level and valence baseline measures

6.10 Appendix 10: Descriptive statistics for main outcomes by high proficiency narrator group

Table 126: Descriptive statistics for the main outcome variables for enthusiastic and calm narrator groups on the higher English proficiency group (LexTALE > 63)

	Group	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>	Skew- ness	Kurto- sis
Narrator emotional tone							
Enthusiasm	Enthusiastic	3.52	1.54	1.00	7.00	−0.10	−0.70
	Calm	1.98	1.32	1.00	6.00	1.30	0.89
Calmness	Enthusiastic	5.38	1.25	2.00	7.00	−0.47	−0.39
	Calm	5.84	1.32	1.00	7.00	−1.63	3.15
Frustration	Enthusiastic	1.38	0.75	1.00	4.00	2.19	4.65
	Calm	1.44	0.93	1.00	6.00	2.91	10.40
Boredom	Enthusiastic	3.35	1.71	1.00	7.00	0.63	−0.39
	Calm	4.54	1.81	1.00	7.00	−0.27	−0.59
Pleasantness	Enthusiastic	4.62	1.16	2.00	7.00	0.26	−0.42
	Calm	4.23	1.52	1.00	7.00	−0.34	−0.84
Activation level	Enthusiastic	3.81	1.47	1.00	7.00	0.04	−0.59
	Calm	2.74	1.54	1.00	6.00	0.58	−0.60
Social partnership with the narrator							
Facilitating learning	Enthusiastic	4.25	0.95	1.70	6.20	−0.43	0.34
	Calm	3.72	1.24	1.00	6.60	0.08	−0.45
Credibility	Enthusiastic	5.26	0.74	3.20	6.60	−0.56	−0.10
	Calm	4.76	1.17	1.20	7.00	−0.87	1.34
Human-like	Enthusiastic	3.87	1.39	1.00	6.80	−0.12	−0.52
	Calm	2.93	1.33	1.00	5.60	0.20	−1.03
Engaging	Enthusiastic	3.14	1.29	1.00	7.00	0.51	0.20
	Calm	2.43	1.35	1.00	5.60	0.81	−0.24
Differences in affective state							
PA change score	Enthusiastic	−0.59	1.17	−4.50	1.75	−0.67	1.63
	Calm	−0.43	0.85	−3.25	2.00	−0.31	1.98

NA change score	Enthusiastic	−0.26	0.98	−3.00	1.25	−0.68	0.46
	Calm	−0.24	0.90	−2.75	1.25	−0.74	0.32
VA change score	Enthusiastic	−0.04	1.01	−2.00	2.00	−0.10	−0.09
	Calm	−0.29	1.04	−2.50	2.50	−0.04	0.09
Activation level change score	Enthusiastic	−0.27	1.63	−4.40	4.20	0.03	1.10
	Calm	−0.36	1.45	−4.40	2.80	−0.41	0.58
Valence change score	Enthusiastic	−0.52	1.35	−3.00	3.60	0.54	0.79
	Calm	−0.70	1.61	−4.40	3.00	−0.21	−0.10
Interest and motivation							
Situational interest	Enthusiastic	3.53	1.12	1.00	5.67	−0.31	−0.69
	Calm	3.38	1.30	1.00	6.17	−0.09	−0.46
Interest (delayed)	Enthusiastic	4.00	1.34	1.00	7.00	−0.27	0.74
	Calm	3.22	1.31	1.00	5.00	−0.28	−1.04
Intrinsic motivation	Enthusiastic	3.50	1.07	1.00	5.75	−0.22	−0.13
	Calm	3.17	1.30	1.00	5.63	0.06	−0.94
Learners' experience							
Paying attention	Enthusiastic	3.35	1.27	1.00	6.00	0.15	−0.24
	Calm	3.34	1.73	1.00	7.00	0.30	−1.04
Difficulty	Enthusiastic	3.25	1.28	1.00	6.00	0.15	−0.60
	Calm	3.05	1.26	1.00	6.00	0.12	−0.74
Exerting more effort	Enthusiastic	3.33	1.28	1.00	6.00	−0.06	−1.05
	Calm	3.39	1.49	1.00	6.00	−0.06	−0.99
Enjoyment	Enthusiastic	3.69	1.32	1.00	6.00	−0.25	−0.52
	Calm	3.59	1.63	1.00	7.00	0.20	−0.81
More lessons like this	Enthusiastic	3.46	1.35	1.00	5.00	−0.61	−0.80
	Calm	3.00	1.63	1.00	6.00	0.34	−0.92
Cognitive outcomes							
Intrinsic cognitive load	Enthusiastic	3.66	1.20	2.00	6.50	0.55	−0.18
	Calm	3.89	1.37	1.00	7.00	0.06	−0.71
Extraneous cognitive load	Enthusiastic	3.43	1.08	1.33	6.67	0.43	0.60
	Calm	3.73	1.35	1.00	7.00	0.22	−0.08

Germane cognitive load	Enthusiastic	4.46	1.05	2.00	6.50	-0.47	-0.08
	Calm	4.45	0.99	2.00	7.00	-0.24	0.45
Mental effort (average)	Enthusiastic	4.44	1.49	1.20	8.60	-0.03	0.38
	Calm	4.51	1.38	1.20	7.40	-0.41	0.18
Immediate testing							
Knowledge	Enthusiastic	13.81	4.39	5.00	23.00	0.15	-0.58
	Calm	15.55	5.16	8.00	26.00	0.46	-0.84
Retention	Enthusiastic	9.23	3.03	3.00	16.00	0.09	-0.58
	Calm	10.13	3.73	4.00	17.00	0.32	-1.09
Transfer	Enthusiastic	4.58	1.89	1.00	8.00	0.19	-0.68
	Calm	5.42	1.89	1.00	10.00	0.44	-0.09
Certainty	Enthusiastic	54.56	17.97	9.21	86.38	-0.42	-0.33
	Calm	56.47	22.78	0.34	95.17	-0.51	-0.13
Certainty in correct answers	Enthusiastic	57.34	20.11	12.60	90.24	-0.33	-0.84
	Calm	59.70	23.46	0.00	95.91	-0.70	0.03
Certainty in incorrect answers	Enthusiastic	50.57	16.21	5.57	79.21	-0.57	0.31
	Calm	51.18	22.09	0.48	93.25	-0.09	-0.19
R Certainty	Enthusiastic	54.25	18.47	8.21	87.63	-0.40	-0.33
	Calm	56.91	22.82	0.53	97.37	-0.53	-0.10
R Certainty in correct answers	Enthusiastic	57.42	20.74	13.30	91.15	-0.29	-0.86
	Calm	60.82	23.77	0.00	100.00	-0.70	0.04
R Certainty in incorrect answers	Enthusiastic	50.02	17.22	2.56	80.00	-0.49	0.19
	Calm	49.97	21.31	0.83	91.00	-0.09	-0.15
T Certainty	Enthusiastic	55.14	19.17	11.10	91.50	-0.33	-0.68
	Calm	55.61	23.49	0.00	92.60	-0.42	-0.36
T Certainty in correct answers	Enthusiastic	58.48	23.08	11.20	100.00	-0.19	-0.99
	Calm	57.73	24.13	0.00	100.00	-0.53	-0.36
T Certainty in incorrect answers	Enthusiastic	51.79	17.74	11.00	80.00	-0.32	-0.63
	Calm	52.79	25.63	0.00	100.00	0.03	-0.50
Self-evaluation	Enthusiastic	3.27	1.27	1.00	6.00	-0.41	-0.14

	Calm	3.42	1.24	1.00	7.00	−0.08	0.62
Delayed testing							
Knowledge	Enthusiastic	12.48	3.39	6.00	21.00	0.21	1.02
	Calm	13.72	4.79	5.00	27.00	0.99	2.79
Retention	Enthusiastic	8.38	2.38	4.00	15.00	0.73	2.00
	Calm	8.83	3.28	3.00	18.00	1.00	2.73
Transfer	Enthusiastic	4.10	1.26	2.00	6.00	−0.36	−0.97
	Calm	4.89	1.88	2.00	9.00	0.48	−0.08
Certainty	Enthusiastic	47.66	18.54	16.55	75.62	−0.10	−1.31
	Calm	41.67	11.58	21.38	66.38	0.61	−0.10
Certainty in correct answers	Enthusiastic	50.81	19.51	18.57	82.56	−0.07	−1.23
	Calm	41.74	12.49	22.00	66.48	0.33	−0.52
Certainty in incorrect answers	Enthusiastic	45.19	17.99	14.67	75.00	−0.11	−1.06
	Calm	40.98	12.50	21.25	65.00	0.76	−0.22
R Certainty	Enthusiastic	44.84	18.67	15.79	73.21	−0.01	−1.37
	Calm	40.40	12.13	17.37	65.00	0.35	−0.26
R Certainty in correct answers	Enthusiastic	48.24	19.27	16.67	83.67	0.03	−1.03
	Calm	42.39	13.74	21.25	74.29	0.55	0.13
R Certainty in incorrect answers	Enthusiastic	41.73	18.55	10.00	72.22	0.01	−1.29
	Calm	37.82	12.20	16.33	63.64	0.54	−0.09
T Certainty	Enthusiastic	53.01	20.49	18.00	82.50	−0.12	−1.25
	Calm	44.07	12.21	29.00	69.00	0.77	−0.11
T Certainty in correct answers	Enthusiastic	56.40	23.13	22.00	93.75	−0.07	−1.40
	Calm	39.87	13.05	25.00	67.78	0.72	−0.28
T Certainty in incorrect answers	Enthusiastic	51.20	20.68	14.00	90.00	0.01	−0.74
	Calm	45.62	14.18	29.00	80.00	0.91	0.22
Self-evaluation	Enthusiastic	3.10	1.34	1.00	5.00	−0.61	−1.32
	Calm	2.56	1.20	1.00	5.00	0.54	−0.71

Note. PA – positive activation, NA – negative activation, VA – valence, R – retention, T – transfer

6.11 Appendix 11: ANCOVA comparisons by high proficiency narrator group

Table 127: ANCOVA comparisons between the enthusiastic and calm narrator on all main dependable variables on the higher English proficiency group (LexTALE > 63)

	ANCOVA*			Homogeneity test**		Normality test	
	<i>F</i>	<i>p</i>	η^2p	<i>F</i>	<i>p</i>	<i>W</i>	<i>p</i>
Narrator emotional tone							
Enthusiasm ^Q	24.534	< .001		5.266	0.024	0.983	0.176
Calmness	3.212	0.076	0.031	0.754	0.387	0.927	< .001
Frustration	0.260	0.611	0.003	0.579	0.449	0.704	< .001
Boredom	10.462	0.002	0.094	0.021	0.886	0.994	0.927
Pleasantness ^Q	1.846	0.177		5.512	0.021	0.991	0.699
Activation level	13.182	< .001	0.115	0.012	0.913	0.990	0.603
Social partnership with the narrator							
Facilitating learning	4.274	0.041	0.041	1.580	0.212	0.994	0.897
Credibility ^Q	6.008	0.016		5.794	0.018	0.948	< .001
Human-like	9.221	0.003	0.084	0.033	0.856	0.984	0.200
Engaging	4.631	0.034	0.044	0.018	0.894	0.967	0.008
Differences in affective state							
Positive activation	2.479	0.119	0.024	1.050	0.308	0.991	0.695
Negative activation	0.115	0.736	0.001	0.001	0.976	0.968	0.009
Valence	0.115	0.735	0.001	0.175	0.677	0.984	0.214
Activation level	0.065	0.799	0.001	0.012	0.914	0.986	0.337
Valence	0.054	0.817	0.001	0.871	0.353	0.989	0.505
Interest and motivation							
Situational interest	0.001	0.979	0.000	0.049	0.825	0.994	0.902
Interest (delayed)	2.616	0.116	0.078	0.063	0.803	0.980	0.715
Intrinsic motivation	0.673	0.414	0.007	0.651	0.422	0.987	0.361
Learners' experience							
Paying attention	0.223	0.638	0.002	3.613	0.060	0.992	0.780

Difficulty	0.296	0.587	0.003	1.247	0.267	0.988	0.433
Exerting more effort	0.001	0.974	0.000	0.572	0.451	0.983	0.179
Enjoyment	0.116	0.734	0.001	0.000	0.995	0.996	0.982
More lessons like this	0.944	0.334	0.009	0.922	0.339	0.983	0.191
Cognitive outcomes							
Intrinsic cognitive load	0.701	0.405	0.007	1.806	0.182	0.989	0.564
Extraneous cognitive load	1.461	0.230	0.014	2.527	0.115	0.982	0.143
Germane cognitive load	0.149	0.701	0.001	0.004	0.951	0.982	0.153
Mental effort (average)	0.000	0.989	0.000	0.606	0.438	0.982	0.156
Immediate testing							
Knowledge	4.416	0.038	0.043	0.690	0.408	0.993	0.893
Retention	1.947	0.166	0.019	1.657	0.201	0.991	0.717
Transfer	6.266	0.014	0.060	0.742	0.391	0.994	0.938
Certainty	0.005	0.945	0.000	1.821	0.180	0.980	0.102
Certainty in correct answers	0.001	0.981	0.000	0.858	0.356	0.975	0.039
Certainty in incorrect answers	0.005	0.945	0.000	2.884	0.092	0.991	0.671
R Certainty	0.089	0.767	0.001	1.628	0.205	0.978	0.071
R Certainty in correct answers	0.036	0.851	0.000	0.576	0.449	0.967	0.010
R Certainty in incorrect answers	0.008	0.928	0.000	0.690	0.408	0.983	0.201
T Certainty	0.121	0.729	0.001	1.056	0.307	0.984	0.219
T Certainty in correct answers	0.156	0.694	0.002	0.001	0.972	0.991	0.691
T Certainty in incorrect answers ^Q	0.020	0.888		4.239	0.042	0.992	0.809
Self-evaluation	0.177	0.675	0.002	3.648	0.059	0.987	0.396
Delayed testing [†]							
Knowledge	0.914	0.346	0.029	1.935	0.173	0.966	0.284
Retention	0.121	0.731	0.004	0.704	0.407	0.972	0.425
Transfer	3.086	0.089	0.091	2.418	0.128	0.971	0.406

Certainty	2.151	0.151	0.058	2.451	0.125	0.976	0.504
Certainty in correct answers	3.375	0.075	0.088	3.247	0.079	0.981	0.694
Certainty in incorrect answers	1.352	0.253	0.037	0.749	0.392	0.983	0.776
R Certainty	1.395	0.245	0.038	1.320	0.257	0.959	0.131
R Certainty in correct answers	1.858	0.182	0.050	0.467	0.498	0.973	0.406
R Certainty in incorrect answers	1.186	0.284	0.033	1.209	0.278	0.970	0.322
T Certainty ^Q	2.959	0.093		4.786	0.034	0.988	0.919
T Certainty in correct answers ^Q	1.429	0.239		10.88 5	0.002	0.986	0.878
T Certainty in incorrect answers	1.449	0.237	0.040	0.363	0.550	0.983	0.784
Self-evaluation	0.954	0.336	0.030	0.565	0.457	0.941	0.040

Note. * $df_1 = 1$, $df_2 = 101$; ** $df_1 = 1$, $df_2 = 107$; † * $df_1 = 1$, $df_2 = 35$; ** $df_1 = 1$, $df_2 = 41$; R – retention, T – transfer; ^Q – Quade test results reported instead of ANCOVA due to the homogeneity of variances assumption not being met; covariates included are prior tested knowledge, prior interest, LexTALE score, and PANAVA-KS baseline measures, except for the Activation level and Valence variables, which had the activation level and valence baseline measures

6.12 Appendix 12: Descriptive statistics for main outcomes by low proficiency SLS group

Table 128: Descriptive statistics for the main outcome variables for the groups without and with SLS on the lower English proficiency group (LexTALE < 63)

	Group	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>	Skew- ness	Kurto- sis
Narrator emotional tone							
Enthusiasm	No SLS	3.00	1.83	1.00	7.00	0.32	−1.07
	SLS	2.83	1.73	1.00	7.00	0.44	−0.94
Calmness	No SLS	5.88	1.29	3.00	7.00	−0.87	−0.39
	SLS	5.60	1.40	2.00	7.00	−0.40	−1.09
Frustration	No SLS	1.56	1.00	1.00	4.00	1.63	1.24
	SLS	1.79	1.37	1.00	6.00	1.61	1.55
Boredom	No SLS	3.58	1.90	1.00	7.00	0.15	−0.91
	SLS	3.55	1.99	1.00	7.00	0.40	−0.92
Pleasantness	No SLS	4.46	1.59	1.00	7.00	−0.24	−0.47
	SLS	4.26	1.43	1.00	7.00	−0.33	−0.47
Activation level	No SLS	3.64	1.40	1.00	7.00	0.12	0.02
	SLS	3.41	1.32	1.00	6.00	−0.11	−0.63
Social partnership with the narrator							
Facilitating learning	No SLS	4.22	1.06	1.30	6.10	−0.75	0.89
	SLS	3.84	1.04	1.20	6.10	−0.20	−0.07
Credibility	No SLS	5.26	1.03	2.60	7.00	−0.61	0.16
	SLS	4.88	1.12	1.00	7.00	−0.96	1.67
Human-like	No SLS	4.38	1.38	1.00	6.40	−0.64	−0.37
	SLS	3.93	1.45	1.00	6.80	−0.19	−0.72
Engaging	No SLS	3.48	1.40	1.00	6.00	0.17	−0.77
	SLS	3.01	1.36	1.00	6.80	0.44	−0.02
Differences in affective state							
PA change score	No SLS	−0.50	1.19	−4.50	2.25	−0.48	1.91
	SLS	−0.59	1.20	−4.00	2.00	−0.52	0.68

NA change score	No SLS	−0.21	1.08	−3.25	2.25	−0.21	0.69
	SLS	−0.22	0.92	−2.75	1.50	−0.62	0.43
VA change score	No SLS	−0.14	1.23	−4.00	3.00	−0.44	1.70
	SLS	−0.46	1.25	−4.00	2.50	−0.24	0.59
Activation level change score	No SLS	−0.22	1.57	−3.80	4.00	0.30	0.66
	SLS	−0.30	1.24	−3.40	3.80	0.12	1.55
Valence change score	No SLS	−0.29	1.54	−4.40	3.40	−0.32	0.64
	SLS	−0.54	1.64	−5.60	2.60	−0.70	0.60
Interest and motivation							
Situational interest	No SLS	3.71	1.03	1.67	5.67	−0.27	−0.68
	SLS	3.20	1.26	1.00	5.83	−0.05	−0.51
Interest (delayed)	No SLS	3.70	0.99	1.00	5.00	−0.62	0.66
	SLS	3.39	1.34	1.00	6.00	0.10	−1.17
Intrinsic motivation	No SLS	3.78	1.09	1.25	6.13	−0.25	−0.29
	SLS	3.33	1.33	1.00	5.88	−0.16	−0.84
Learners' experience							
Paying attention	No SLS	3.49	1.29	1.00	6.00	0.19	−0.57
	SLS	3.26	1.48	1.00	6.00	0.04	−1.18
Difficulty	No SLS	3.88	1.40	1.00	7.00	0.02	−0.68
	SLS	3.98	1.56	1.00	7.00	0.03	−0.66
Exerting more effort	No SLS	3.85	1.51	1.00	7.00	−0.17	−0.94
	SLS	3.69	1.40	1.00	7.00	0.03	−0.23
Enjoyment	No SLS	3.73	1.34	1.00	7.00	0.21	−0.10
	SLS	3.47	1.58	1.00	6.00	−0.29	−1.06
More lessons like this	No SLS	3.69	1.66	1.00	7.00	0.30	−0.61
	SLS	3.03	1.47	1.00	6.00	0.04	−0.83
Cognitive outcomes							
Intrinsic cognitive load	No SLS	4.23	1.11	1.50	6.00	−0.27	−0.38
	SLS	4.17	1.24	1.50	6.50	−0.13	−0.63
Extraneous cognitive load	No SLS	3.86	1.13	1.67	6.00	0.10	−0.89
	SLS	3.64	1.34	1.33	7.00	0.56	−0.34

Germane cognitive load	No SLS	4.77	0.97	2.50	7.00	-0.27	-0.22
	SLS	4.35	1.39	1.00	6.50	-0.53	-0.68
Mental effort (average)	No SLS	4.97	1.48	1.40	8.80	-0.21	0.56
	SLS	4.61	1.34	1.80	7.60	-0.16	-0.62
Immediate testing							
Knowledge	No SLS	12.25	3.52	5.00	23.00	0.53	0.51
	SLS	12.38	4.15	5.00	24.00	0.71	0.39
Retention	No SLS	8.17	2.54	4.00	16.00	0.61	0.67
	SLS	8.17	2.88	2.00	15.00	0.33	0.18
Transfer	No SLS	4.08	1.86	0.00	8.00	-0.19	-0.32
	SLS	4.21	2.01	0.00	9.00	0.23	-0.42
Certainty	No SLS	48.94	19.50	6.66	88.59	-0.14	-0.20
	SLS	48.68	19.18	8.62	84.83	-0.17	-0.85
Certainty in correct answers	No SLS	53.44	21.86	6.15	93.33	-0.16	-0.56
	SLS	50.66	19.43	10.00	87.92	-0.03	-0.85
Certainty in incorrect answers	No SLS	45.57	18.61	3.07	84.61	-0.20	-0.11
	SLS	46.84	19.06	0.00	79.08	-0.38	-0.61
R Certainty	No SLS	48.67	20.16	6.47	90.00	-0.02	-0.45
	SLS	48.09	18.76	9.21	82.11	-0.14	-1.00
R Certainty in correct answers	No SLS	53.13	22.19	6.11	96.43	-0.11	-0.79
	SLS	50.27	19.76	12.50	93.18	0.12	-0.83
R Certainty in incorrect answers	No SLS	45.07	19.12	2.00	85.56	-0.04	-0.42
	SLS	45.90	18.93	0.00	85.57	-0.32	-0.59
T Certainty	No SLS	49.44	21.96	0.00	92.50	-0.21	-0.28
	SLS	49.79	21.47	0.00	90.00	-0.15	-0.66
T Certainty in correct answers	No SLS	52.48	24.38	0.00	100.00	-0.11	-0.58
	SLS	50.90	23.70	0.00	97.50	-0.01	-0.81
T Certainty in incorrect answers	No SLS	47.34	21.45	0.00	91.25	-0.25	-0.28
	SLS	49.20	22.51	0.00	100.00	0.14	-0.29
Self-evaluation	No SLS	3.42	1.10	1.00	7.00	0.20	1.53

	SLS	3.21	1.09	1.00	6.00	−0.09	0.31
Delayed testing							
Knowledge	No SLS	10.96	3.40	4.00	21.00	0.50	1.89
	SLS	11.82	4.28	4.00	20.00	−0.02	−0.92
Retention	No SLS	7.30	2.61	2.00	14.00	0.28	0.56
	SLS	7.64	2.91	3.00	14.00	0.30	−0.75
Transfer	No SLS	3.67	1.27	2.00	7.00	0.45	0.15
	SLS	4.18	2.07	0.00	7.00	−0.34	−0.92
Certainty	No SLS	45.74	18.38	7.38	75.62	−0.18	−0.78
	SLS	37.85	22.36	0.00	79.31	−0.15	−0.59
Certainty in correct answers	No SLS	47.48	19.10	7.06	82.56	−0.16	−0.62
	SLS	41.18	22.78	0.00	78.89	−0.37	−0.67
Certainty in incorrect answers	No SLS	44.96	18.57	7.77	75.00	−0.15	−0.97
	SLS	35.22	22.05	0.00	80.36	0.04	−0.37
R Certainty	No SLS	43.89	17.98	5.79	73.21	−0.08	−0.70
	SLS	35.67	21.76	0.00	76.32	−0.05	−0.65
R Certainty in correct answers	No SLS	44.51	18.67	7.82	83.67	0.20	−0.39
	SLS	39.56	22.93	0.00	78.89	−0.24	−0.91
R Certainty in incorrect answers	No SLS	43.26	18.48	3.00	72.22	−0.17	−0.85
	SLS	32.81	20.97	0.00	77.78	0.16	−0.27
T Certainty	No SLS	49.25	20.29	10.40	82.50	−0.23	−1.12
	SLS	41.99	24.14	0.00	85.00	−0.23	−0.54
T Certainty in correct answers	No SLS	52.10	23.20	5.40	85.00	−0.26	−1.14
	SLS	42.12	23.28	0.00	85.00	−0.44	−0.44
T Certainty in incorrect answers	No SLS	47.94	19.65	14.00	80.14	−0.08	−1.15
	SLS	39.03	25.17	0.00	85.00	0.15	−0.67
Self-evaluation	No SLS	3.22	0.93	1.00	5.00	−0.48	−0.22
	SLS	3.18	0.98	1.00	5.00	−0.64	0.04

Note. PA – positive activation, NA – negative activation, VA – valence, R – retention, T – transfer

6.13 Appendix 13: ANCOVA comparisons by low proficiency SLS group

Table 129: ANCOVA comparisons between the groups without and with SLS on all main dependable variables on the lower English proficiency group (LexTALE < 63)

	ANCOVA*			Homogeneity test**		Normality test	
	<i>F</i>	<i>p</i>	η^2p	<i>F</i>	<i>p</i>	<i>W</i>	<i>p</i>
Narrator emotional tone							
Enthusiasm	0.185	0.668	0.002	0.983	0.324	0.910	< .001
Calmness	0.290	0.591	0.003	0.269	0.605	0.935	< .001
Frustration	0.263	0.609	0.002	3.197	0.076	0.859	< .001
Boredom	0.161	0.689	0.001	0.015	0.904	0.958	0.001
Pleasantness	0.025	0.876	0.000	3.300	0.072	0.980	0.075
Activation level	0.031	0.860	0.000	0.865	0.354	0.989	0.448
Social partnership with the narrator							
Facilitating learning	1.973	0.163	0.018	0.000	0.996	0.983	0.134
Credibility	1.660	0.200	0.015	0.869	0.353	0.966	0.004
Human-like	2.934	0.090	0.026	0.213	0.645	0.984	0.167
Engaging	2.561	0.112	0.023	0.492	0.485	0.986	0.267
Differences in affective state							
Positive activation	0.788	0.377	0.007	0.023	0.881	0.984	0.168
Negative activation	0.186	0.667	0.002	1.138	0.288	0.994	0.925
Valence	0.810	0.370	0.007	0.224	0.637	0.977	0.043
Activation level	1.209	0.274	0.011	1.146	0.287	0.983	0.140
Valence	0.855	0.357	0.008	0.088	0.767	0.951	< .001
Interest and motivation							
Situational interest ^Q	3.486	0.064		4.527	0.035	0.991	0.650
Interest (delayed)	0.036	0.850	0.001	0.050	0.824	0.980	0.497
Intrinsic motivation	1.718	0.193	0.016	0.799	0.373	0.991	0.665
Learners' experience							
Paying attention ^Q	0.046	0.831		3.899	0.051	0.989	0.449
Difficulty	0.093	0.761	0.001	0.987	0.323	0.985	0.212

Exerting more effort	0.046	0.830	0.000	0.911	0.342	0.986	0.247
Enjoyment	0.204	0.652	0.002	1.454	0.230	0.995	0.932
More lessons like this	3.116	0.080	0.028	2.046	0.155	0.980	0.078
Cognitive outcomes							
Intrinsic cognitive load	0.004	0.949	0.000	2.939	0.089	0.987	0.307
Extraneous cognitive load	1.516	0.221	0.014	2.241	0.137	0.975	0.026
Germane cognitive load ^Q	0.788	0.377		7.013	0.009	0.974	0.024
Mental effort (average)	0.707	0.402	0.006	0.000	0.995	0.981	0.092
Immediate testing							
Knowledge	0.026	0.871	0.000	2.686	0.104	0.982	0.128
Retention	0.002	0.963	0.000	0.172	0.679	0.985	0.239
Transfer	0.060	0.807	0.001	1.812	0.181	0.992	0.717
Certainty	0.004	0.952	0.000	0.033	0.857	0.983	0.140
Certainty in correct answers	0.499	0.482	0.005	1.517	0.221	0.986	0.273
Certainty in incorrect answers	0.156	0.694	0.001	0.045	0.833	0.983	0.156
R Certainty	0.002	0.962	0.000	0.158	0.692	0.988	0.370
R Certainty in correct answers	0.405	0.526	0.004	1.255	0.265	0.989	0.486
R Certainty in incorrect answers	0.134	0.715	0.001	0.015	0.903	0.987	0.341
T Certainty	0.005	0.942	0.000	0.151	0.698	0.983	0.135
T Certainty in correct answers	0.388	0.534	0.004	0.393	0.532	0.985	0.220
T Certainty in incorrect answers	0.119	0.731	0.001	0.015	0.902	0.991	0.641
Self-evaluation	0.199	0.656	0.002	0.033	0.855	0.979	0.064
Delayed testing [†]							
Knowledge	2.115	0.153	0.043	3.011	0.089	0.972	0.218
Retention	0.978	0.328	0.020	2.123	0.151	0.961	0.070
Transfer ^Q	3.595	0.063		4.904	0.031	0.990	0.930

Certainty	2.755	0.104	0.058	0.191	0.664	0.983	0.630
Certainty in correct answers	1.596	0.213	0.034	0.511	0.478	0.986	0.807
Certainty in incorrect answers	3.947	0.053	0.081	0.036	0.849	0.981	0.578
R Certainty	3.030	0.089	0.063	0.085	0.772	0.986	0.774
R Certainty in correct answers	1.215	0.276	0.026	1.029	0.315	0.985	0.759
R Certainty in incorrect answers	4.691	0.036	0.094	0.027	0.869	0.988	0.884
T Certainty	2.129	0.151	0.045	0.054	0.817	0.980	0.530
T Certainty in correct answers	2.444	0.125	0.053	0.043	0.837	0.969	0.184
T Certainty in incorrect answers	2.566	0.116	0.054	0.596	0.444	0.982	0.623
Self-evaluation	0.025	0.876	0.001	0.000	0.996	0.983	0.629

Note. * $df_1 = 1$, $df_2 = 109$; ** $df_1 = 1$, $df_2 = 115$; † * $df_1 = 1$, $df_2 = 47$; ** $df_1 = 1$, $df_2 = 53$; R – retention, T – transfer, ^Q – Quade test results reported instead of ANCOVA due to the homogeneity of variances assumption not being met; covariates included are prior tested knowledge, prior interest, LexTALE score, and PANAVA-KS baseline measures, except for the Activation level and Valence variables, which had the activation level and valence baseline measures

6.14 Appendix 14: Descriptive statistics for main outcomes by high proficiency SLS group

Table 130: Descriptive statistics for the main outcome variables for the groups without and with SLS on the higher English proficiency group (LexTALE > 63)

	Group	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>	Skew- ness	Kurto- sis
Narrator emotional tone							
Enthusiasm	No SLS	2.88	1.68	1.00	7.00	0.44	−0.92
	SLS	2.55	1.54	1.00	6.00	0.55	−0.90
Calmness	No SLS	5.46	1.35	1.00	7.00	−0.91	0.86
	SLS	5.79	1.25	2.00	7.00	−1.20	1.45
Frustration	No SLS	1.50	0.95	1.00	6.00	2.61	8.55
	SLS	1.32	0.70	1.00	4.00	2.59	6.96
Boredom	No SLS	3.77	1.89	1.00	7.00	0.40	−0.81
	SLS	4.19	1.82	1.00	7.00	−0.11	−0.80
Pleasantness	No SLS	4.18	1.38	1.00	6.00	−0.33	−0.74
	SLS	4.66	1.33	1.00	7.00	−0.21	−0.11
Activation level	No SLS	3.39	1.63	1.00	6.00	0.07	−0.96
	SLS	3.09	1.56	1.00	7.00	0.44	−0.54
Social partnership with the narrator							
Facilitating learning	No SLS	3.95	1.21	1.00	6.20	−0.38	−0.31
	SLS	4.00	1.06	1.90	6.60	0.02	−0.31
Credibility	No SLS	4.99	0.99	1.40	6.60	−1.27	2.56
	SLS	5.01	1.05	1.20	7.00	−0.94	2.04
Human-like	No SLS	3.51	1.49	1.00	6.80	−0.12	−0.81
	SLS	3.23	1.36	1.00	6.20	0.27	−0.68
Engaging	No SLS	2.78	1.52	1.00	7.00	0.63	−0.37
	SLS	2.76	1.19	1.00	5.60	0.41	−0.50
Differences in affective state							
PA change score	No SLS	−0.51	1.08	−3.25	2.00	0.05	0.49
	SLS	−0.50	0.95	−4.50	1.25	−1.73	5.26

NA change score	No SLS	−0.18	0.97	−3.00	1.25	−0.73	0.47
	SLS	−0.33	0.89	−2.75	1.25	−0.75	0.36
VA change score	No SLS	−0.21	1.13	−2.50	2.00	0.11	−0.45
	SLS	−0.12	0.92	−2.00	2.50	−0.36	0.77
Activation level change score	No SLS	−0.25	1.59	−4.40	4.20	0.03	1.48
	SLS	−0.38	1.48	−4.40	2.60	−0.42	0.10
Valence change score	No SLS	−0.58	1.52	−4.40	3.60	0.10	0.88
	SLS	−0.65	1.48	−3.60	3.00	−0.08	−0.27
Interest and motivation							
Situational interest	No SLS	3.45	1.18	1.00	6.00	−0.31	−0.45
	SLS	3.45	1.25	1.00	6.17	−0.10	−0.55
Interest (delayed)	No SLS	4.05	1.20	2.00	7.00	0.09	0.88
	SLS	3.17	1.42	1.00	5.00	−0.19	−1.19
Intrinsic motivation	No SLS	3.28	1.25	1.00	5.63	−0.17	−0.70
	SLS	3.37	1.16	1.00	5.75	−0.03	−0.64
Learners' experience							
Paying attention	No SLS	3.33	1.55	1.00	6.00	−0.17	−1.16
	SLS	3.11	1.48	1.00	6.00	0.02	−0.86
Difficulty	No SLS	3.64	1.51	1.00	7.00	0.08	−0.70
	SLS	3.64	1.47	1.00	7.00	−0.02	−0.61
Exerting more effort	No SLS	3.36	1.43	1.00	6.00	0.03	−1.05
	SLS	3.36	1.35	1.00	6.00	−0.15	−0.85
Enjoyment	No SLS	3.13	1.31	1.00	6.00	0.12	−0.49
	SLS	3.17	1.24	1.00	6.00	0.17	−0.89
More lessons like this	No SLS	3.18	1.44	1.00	6.00	0.40	−0.36
	SLS	3.51	1.59	1.00	7.00	0.12	−0.87
Cognitive outcomes							
Intrinsic cognitive load	No SLS	3.70	1.22	1.00	6.50	0.37	−0.33
	SLS	3.87	1.37	1.50	7.00	0.18	−0.72
Extraneous cognitive load	No SLS	3.78	1.30	1.33	7.00	0.47	−0.16
	SLS	3.39	1.13	1.00	6.67	0.06	0.36

Germane cognitive load	No SLS	4.44	0.94	2.00	6.00	-0.45	-0.25
	SLS	4.47	1.09	2.00	7.00	-0.30	0.31
Mental effort (average)	No SLS	4.43	1.39	1.60	8.60	0.34	0.83
	SLS	4.52	1.48	1.20	7.00	-0.71	-0.05
Immediate testing							
Knowledge	No SLS	14.11	4.24	8.00	26.00	1.05	0.80
	SLS	15.33	5.41	5.00	26.00	-0.04	-0.94
Retention	No SLS	9.24	2.90	4.00	16.00	0.80	0.07
	SLS	10.17	3.87	3.00	17.00	-0.06	-1.06
Transfer	No SLS	4.87	1.86	1.00	10.00	0.59	0.47
	SLS	5.15	2.01	1.00	9.00	0.02	-0.67
Certainty	No SLS	54.11	21.09	0.34	91.72	-0.47	-0.14
	SLS	57.05	19.96	3.45	95.17	-0.42	-0.02
Certainty in correct answers	No SLS	57.58	21.86	0.00	91.48	-0.56	-0.11
	SLS	59.59	21.95	2.50	95.91	-0.52	-0.44
Certainty in incorrect answers	No SLS	50.07	20.77	0.48	93.25	-0.23	-0.17
	SLS	51.74	17.93	3.81	92.86	-0.13	0.64
R Certainty	No SLS	54.13	21.18	0.53	91.26	-0.49	-0.11
	SLS	57.20	20.41	4.74	97.37	-0.40	-0.15
R Certainty in correct answers	No SLS	58.29	22.09	0.00	94.00	-0.56	-0.07
	SLS	60.09	22.72	5.00	100.00	-0.46	-0.58
R Certainty in incorrect answers	No SLS	49.78	20.94	0.83	91.00	-0.23	-0.09
	SLS	50.22	17.69	4.67	87.50	-0.21	0.17
T Certainty	No SLS	54.09	22.36	0.00	92.60	-0.34	-0.54
	SLS	56.76	20.46	1.00	92.00	-0.42	-0.14
T Certainty in correct answers	No SLS	56.95	24.55	0.00	100.00	-0.27	-0.72
	SLS	59.31	22.56	0.00	100.00	-0.50	-0.49
T Certainty in incorrect answers	No SLS	50.87	22.61	0.00	100.00	-0.16	-0.43
	SLS	53.79	21.50	1.67	100.00	0.13	0.01
Self-evaluation	No SLS	3.45	1.21	1.00	7.00	0.14	0.95

	SLS	3.23	1.29	1.00	6.00	−0.56	−0.57
Delayed testing							
Knowledge	No SLS	12.29	2.65	6.00	16.00	−0.64	−0.02
	SLS	13.94	5.24	5.00	27.00	0.67	1.20
Retention	No SLS	8.29	1.93	4.00	11.00	−0.17	−0.37
	SLS	8.94	3.59	3.00	18.00	0.90	1.45
Transfer	No SLS	4.00	1.18	2.00	6.00	0.00	−0.89
	SLS	5.00	1.88	2.00	9.00	0.18	0.10
Certainty	No SLS	46.14	16.67	16.55	75.62	0.33	−0.90
	SLS	42.81	14.22	21.38	66.38	0.14	−1.10
Certainty in correct answers	No SLS	48.51	17.45	18.57	82.56	0.31	−0.66
	SLS	43.47	15.90	22.00	75.00	0.36	−1.02
Certainty in incorrect answers	No SLS	44.31	16.61	14.67	75.00	0.33	−0.91
	SLS	41.57	14.13	16.18	65.00	0.04	−0.83
R Certainty	No SLS	44.29	16.91	15.79	73.21	0.29	−1.11
	SLS	40.59	14.20	15.79	65.00	0.00	−0.83
R Certainty in correct answers	No SLS	47.50	17.24	16.67	83.67	0.27	−0.37
	SLS	42.66	16.18	21.25	74.29	0.43	−0.87
R Certainty in incorrect answers	No SLS	41.32	17.53	15.00	72.22	0.38	−1.35
	SLS	37.90	13.16	10.00	60.42	−0.30	−0.31
T Certainty	No SLS	49.66	17.30	18.00	82.50	0.24	−0.55
	SLS	47.02	17.36	25.00	82.50	0.65	−0.82
T Certainty in correct answers	No SLS	50.24	20.70	22.00	85.00	0.43	−1.14
	SLS	45.30	19.91	25.00	93.75	0.87	0.08
T Certainty in incorrect answers	No SLS	48.92	16.57	14.00	80.14	0.07	−0.10
	SLS	47.69	19.26	25.00	90.00	0.71	−0.45
Self-evaluation	No SLS	2.90	1.22	1.00	5.00	0.02	−1.55
	SLS	2.78	1.40	1.00	5.00	−0.14	−1.60

Note. PA – positive activation, NA – negative activation, VA – valence, R – retention, T – transfer

6.15 Appendix 15: ANCOVA comparisons by high proficiency SLS group

Table 131: ANCOVA comparisons between the groups without and with SLS on all main dependable variables on the higher English proficiency group (LexTALE > 63)

	ANCOVA*			Homogeneity test**		Normality test	
	<i>F</i>	<i>p</i>	η^2p	<i>F</i>	<i>p</i>	<i>W</i>	<i>p</i>
Narrator emotional tone							
Enthusiasm	0.940	0.334	0.009	1.070	0.303	0.943	< .001
Calmness	1.776	0.186	0.017	1.207	0.274	0.924	< .001
Frustration	1.519	0.221	0.015	2.217	0.139	0.715	< .001
Boredom	1.129	0.290	0.011	0.000	0.992	0.983	0.183
Pleasantness	3.582	0.061	0.034	1.074	0.302	0.993	0.880
Activation level	0.925	0.338	0.009	0.054	0.817	0.987	0.351
Social partnership with the narrator							
Facilitating learning	0.130	0.719	0.001	0.061	0.806	0.991	0.721
Credibility	0.242	0.624	0.002	0.102	0.751	0.942	< .001
Human-like	0.733	0.394	0.007	0.413	0.522	0.984	0.200
Engaging	0.001	0.982	0.000	1.891	0.172	0.967	0.009
Differences in affective state							
Positive activation	0.001	0.976	0.000	1.290	0.259	0.989	0.540
Negative activation	0.093	0.761	0.001	0.651	0.422	0.967	0.008
Valence	0.001	0.976	0.000	0.855	0.357	0.983	0.168
Activation level	0.855	0.357	0.008	0.001	0.973	0.987	0.356
Valence	0.084	0.773	0.001	1.706	0.194	0.986	0.312
Interest and motivation							
Situational interest	0.114	0.736	0.001	0.351	0.555	0.994	0.930
Interest (delayed)	3.204	0.083	0.094	1.228	0.275	0.976	0.575
Intrinsic motivation	0.289	0.592	0.003	0.037	0.849	0.988	0.415
Learners' experience							
Paying attention	1.938	0.167	0.019	2.617	0.109	0.995	0.955
Difficulty	0.049	0.826	0.000	0.351	0.555	0.988	0.428

Exerting more effort	0.123	0.727	0.001	0.225	0.636	0.981	0.135
Enjoyment	0.030	0.862	0.000	0.100	0.752	0.995	0.973
More lessons like this	0.468	0.495	0.005	0.244	0.623	0.982	0.167
Cognitive outcomes							
Intrinsic cognitive load	0.503	0.480	0.005	1.891	0.172	0.989	0.516
Extraneous cognitive load	2.451	0.121	0.024	0.910	0.342	0.980	0.109
Germane cognitive load	0.106	0.746	0.001	0.175	0.676	0.982	0.159
Mental effort (average)	0.024	0.877	0.000	0.281	0.597	0.982	0.146
Immediate testing							
Knowledge	2.022	0.158	0.020	1.872	0.174	0.985	0.260
Retention	2.081	0.152	0.021	2.387	0.125	0.988	0.461
Transfer	0.620	0.433	0.006	0.133	0.716	0.992	0.769
Certainty	0.476	0.492	0.005	1.130	0.290	0.979	0.094
Certainty in correct answers	0.103	0.749	0.001	0.318	0.574	0.976	0.049
Certainty in incorrect answers	0.223	0.638	0.002	1.434	0.234	0.990	0.608
R Certainty	0.539	0.464	0.005	1.063	0.305	0.979	0.085
R Certainty in correct answers	0.064	0.801	0.001	0.506	0.478	0.968	0.011
R Certainty in incorrect answers	0.040	0.843	0.000	1.005	0.318	0.984	0.229
T Certainty	0.311	0.578	0.003	1.242	0.268	0.985	0.290
T Certainty in correct answers	0.159	0.691	0.002	1.134	0.289	0.992	0.809
T Certainty in incorrect answers	0.338	0.562	0.003	0.378	0.540	0.990	0.636
Self-evaluation	1.161	0.284	0.012	0.194	0.661	0.985	0.281
Delayed testing [†]							
Knowledge	0.353	0.557	0.011	1.976	0.168	0.976	0.578
Retention	0.002	0.968	0.000	0.270	0.606	0.971	0.393
Transfer	2.343	0.136	0.070	2.996	0.092	0.987	0.933
Certainty	0.106	0.747	0.003	0.091	0.765	0.974	0.415

Certainty in correct answers	0.351	0.558	0.010	0.086	0.771	0.991	0.981
Certainty in incorrect answers	0.026	0.872	0.001	0.092	0.763	0.976	0.495
R Certainty	0.168	0.685	0.005	0.163	0.689	0.965	0.214
R Certainty in correct answers	0.419	0.522	0.012	0.001	0.982	0.985	0.839
R Certainty in incorrect answers	0.034	0.854	0.001	1.055	0.310	0.970	0.318
T Certainty	0.022	0.883	0.001	0.188	0.667	0.979	0.614
T Certainty in correct answers	0.105	0.747	0.003	0.171	0.681	0.962	0.159
T Certainty in incorrect answers	0.105	0.747	0.003	0.171	0.681	0.962	0.159
Self-evaluation	0.003	0.956	0.000	3.345	0.075	0.909	0.004

Note. * $df_1 = 1$, $df_2 = 101$; ** $df_1 = 1$, $df_2 = 107$; † * $df_1 = 1$, $df_2 = 35$; ** $df_1 = 1$, $df_2 = 41$; R – retention, T – transfer; covariates included are prior tested knowledge, prior interest, LexTALE score, and PANAVA-KS baseline measures, except for the Activation level and Valence variables, which had the activation level and valence baseline measures

6.16 Appendix 16: Two-way ANCOVA comparisons for low proficiency group – Study 1

Table 132: Two-way ANCOVA comparisons on all main dependable variables on the lower English proficiency group (LexTALE < 63), together with homogeneity tests

		ANCOVA*			Levene's test**	
		<i>F</i>	<i>p</i>	η^2p	<i>F</i>	<i>p</i>
Narrator affective state						
Enthusiasm	Narrator emotion	18.137	< .001	0.145	0.331	0.803
	SLS	0.525	0.470	0.005		
	Interaction	0.004	0.951	0.000		
Calmness	Narrator emotion	11.858	< .001	0.100	0.193	0.901
	SLS	0.104	0.748	0.001		
	Interaction	0.865	0.354	0.008		
Frustration	Narrator emotion	0.127	0.723	0.001	1.900	0.134
	SLS	0.284	0.595	0.003		
	Interaction	0.021	0.885	0.000		
Boredom	Narrator emotion	10.788	0.001	0.092	0.589	0.623
	SLS	0.038	0.847	0.000		
	Interaction	0.434	0.512	0.004		
Pleasantness	Narrator emotion	5.587	0.020	0.050	0.440	0.725
	SLS	0.060	0.808	0.001		
	Interaction	4.229	0.042	0.038		
Activation level	Narrator emotion	5.866	0.017	0.052	0.780	0.507
	SLS	0.101	0.751	0.001		
	Interaction	0.109	0.742	0.001		
Narrator perception						
Facilitating learning	Narrator emotion	4.611	0.034	0.041	1.560	0.203
	SLS	2.446	0.121	0.022		
	Interaction	0.169	0.682	0.002		
Credibility	Narrator emotion	8.087	0.005	0.070	3.883	0.011
	SLS	2.261	0.136	0.021		
	Interaction	0.054	0.817	0.001		
Human-like	Narrator emotion	4.964	0.028	0.044	1.858	0.141
	SLS	3.385	0.069	0.031		

Engaging	Interaction	2.299	0.132	0.021	0.602	0.615
	Narrator emotion	4.518	0.036	0.041		
	SLS	2.937	0.089	0.027		
	Interaction	2.901	0.091	0.026		
Participants' affective state						
Positive activation	Narrator emotion	0.182	0.671	0.002	0.046	0.987
	SLS	0.732	0.394	0.007		
	Interaction	0.011	0.917	0.000		
Negative activation	Narrator emotion	1.149	0.286	0.011	0.580	0.629
	SLS	0.263	0.609	0.002		
	Interaction	0.317	0.575	0.003		
Valence	Narrator emotion	1.334	0.251	0.012	1.127	0.341
	SLS	0.656	0.420	0.006		
	Interaction	0.245	0.622	0.002		
Activation level [†]	Narrator emotion	1.199	0.276	0.011	1.247	0.296
	SLS	1.825	0.180	0.017		
	Interaction	0.688	0.409	0.006		
Valence [†]	Narrator emotion	0.448	0.505	0.004	0.462	0.709
	SLS	0.088	0.767	0.001		
	Interaction	0.350	0.555	0.003		
Interest and motivation						
Situational interest	Narrator emotion	0.772	0.382	0.007	1.799	0.152
	SLS	2.731	0.101	0.025		
	Interaction	0.117	0.733	0.001		
Interest (delayed) [‡]	Narrator emotion	0.116	0.735	0.003	0.517	0.672
	SLS	0.048	0.827	0.001		
	Interaction	0.138	0.712	0.003		
Intrinsic motivation	Narrator emotion	0.975	0.326	0.009	0.606	0.612
	SLS	1.810	0.181	0.017		
	Interaction	0.389	0.534	0.004		
Learners' experience						
Paying attention	Narrator emotion	0.972	0.326	0.009	1.456	0.230
	SLS	0.115	0.736	0.001		
	Interaction	0.127	0.723	0.001		
Difficulty	Narrator emotion	0.165	0.685	0.002	0.896	0.446

Exerting more effort			SLS	0.085	0.771	0.001	0.813	0.489
			Interaction	0.231	0.632	0.002		
			Narrator emotion	0.208	0.649	0.002		
Enjoyment			SLS	0.049	0.825	0.000	0.447	0.720
			Interaction	0.424	0.516	0.004		
			Narrator emotion	0.125	0.724	0.001		
More lessons like this			SLS	0.242	0.624	0.002	0.796	0.498
			Interaction	0.488	0.486	0.005		
			Narrator emotion	0.096	0.757	0.001		
			SLS	3.075	0.082	0.028		
			Interaction	0.160	0.690	0.001		
Cognitive outcomes								
Intrinsic load	cognitive	Narrator emotion	3.512	0.064	0.032	1.632	0.186	
		SLS	0.007	0.932	0.000			
		Interaction	1.254	0.265	0.012			
Extraneous load	cognitive	Narrator emotion	0.470	0.494	0.004	1.163	0.327	
		SLS	1.413	0.237	0.013			
		Interaction	0.112	0.738	0.001			
Germane load	cognitive	Narrator emotion	0.223	0.638	0.002	2.387	0.073	
		SLS	1.598	0.209	0.015			
		Interaction	0.332	0.566	0.003			
Mental effort average		Narrator emotion	1.699	0.195	0.016	0.363	0.780	
		SLS	0.863	0.355	0.008			
		Interaction	0.110	0.741	0.001			
Immediate testing								
Knowledge		Narrator emotion	0.420	0.518	0.004	1.976	0.122	
		SLS	0.015	0.903	0.000			
		Interaction	0.001	0.976	0.000			
Retention		Narrator emotion	0.670	0.415	0.006	0.276	0.843	
		SLS	0.000	0.990	0.000			
		Interaction	0.079	0.780	0.001			
Transfer		Narrator emotion	0.008	0.930	0.000	1.572	0.200	
		SLS	0.063	0.802	0.001			
		Interaction	0.205	0.651	0.002			
Certainty		Narrator emotion	0.692	0.407	0.006	1.315	0.273	

	SLS	0.016	0.899	0.000		
	Interaction	0.263	0.609	0.002		
Certainty in correct answers	Narrator emotion	0.117	0.733	0.001	2.219	0.090
	SLS	0.547	0.461	0.005		
	Interaction	0.370	0.544	0.003		
Certainty in incorrect answers	Narrator emotion	1.216	0.273	0.011	1.656	0.181
	SLS	0.092	0.762	0.001		
	Interaction	0.511	0.476	0.005		
Self-evaluation	Narrator emotion	2.503	0.117	0.023	1.762	0.159
	SLS	0.134	0.715	0.001		
	Interaction	0.305	0.582	0.003		
Delayed testing [‡]						
Knowledge	Narrator emotion	7.638	0.008	0.145	0.140	0.936
	SLS	1.727	0.196	0.037		
	Interaction	2.479	0.122	0.052		
Retention	Narrator emotion	4.060	0.050	0.083	0.743	0.532
	SLS	0.690	0.411	0.015		
	Interaction	1.133	0.293	0.025		
Transfer	Narrator emotion	7.978	0.007	0.151	0.161	0.922
	SLS	2.458	0.124	0.052		
	Interaction	3.087	0.086	0.064		
Certainty	Narrator emotion	0.387	0.537	0.009	0.111	0.953
	SLS	2.572	0.116	0.056		
	Interaction	0.626	0.433	0.014		
Certainty in correct answers	Narrator emotion	0.854	0.360	0.019	0.250	0.861
	SLS	1.457	0.234	0.033		
	Interaction	0.751	0.391	0.017		
Certainty in incorrect answers	Narrator emotion	0.135	0.715	0.003	0.240	0.868
	SLS	3.728	0.060	0.080		
	Interaction	0.619	0.436	0.014		
Self-evaluation	Narrator emotion	2.277	0.138	0.048	1.980	0.129
	SLS	0.093	0.761	0.002		
	Interaction	0.454	0.504	0.010		

Note. * $df_1 = 1$, $df_2 = 107$; ** $df_1 = 3$, $df_2 = 113$; [†] – * $df_1 = 1$, $df_2 = 108$; ** $df_1 = 3$, $df_2 = 113$; [‡] – * $df_1 = 1$, $df_2 = 43$; ** $df_1 = 3$, $df_2 = 49$; covariates: prior tested knowledge, prior interest, LexTALE, and PANAVA-KS / activation level and valence baseline measures

6.17 Appendix 17: Two-way ANCOVA comparisons for high proficiency group – Study 1

Table 133: Two-way ANCOVA comparisons on all main dependable variables on the higher English proficiency group (LexTALE > 63), together with homogeneity tests

		ANCOVA*			Levene's test**	
		<i>F</i>	<i>p</i>	η^2p	<i>F</i>	<i>p</i>
Narrator affective state						
Enthusiasm	Narrator emotion	23.383	< .001	0.191	2.223	0.090
	SLS	0.495	0.483	0.005		
	Interaction	0.532	0.468	0.005		
Calmness	Narrator emotion	2.890	0.092	0.028	0.890	0.449
	SLS	1.481	0.227	0.015		
	Interaction	0.001	0.981	0.000		
Frustration	Narrator emotion	0.362	0.549	0.004	1.391	0.250
	SLS	1.505	0.223	0.015		
	Interaction	0.679	0.412	0.007		
Boredom	Narrator emotion	9.898	0.002	0.091	0.739	0.531
	SLS	0.771	0.382	0.008		
	Interaction	0.027	0.870	0.000		
Pleasantness	Narrator emotion	3.619	0.060	0.035	3.076	0.031
	SLS	3.966	0.049	0.039		
	Interaction	1.325	0.253	0.013		
Activation level	Narrator emotion	12.709	< .001	0.114	0.282	0.838
	SLS	0.511	0.476	0.005		
	Interaction	1.675	0.199	0.017		
Narrator perception						
Facilitating learning	Narrator emotion	4.368	0.039	0.042	0.963	0.413
	SLS	0.225	0.636	0.002		
	Interaction	0.534	0.467	0.005		
Credibility	Narrator emotion	7.256	0.008	0.068	2.093	0.106
	SLS	0.417	0.520	0.004		
	Interaction	0.935	0.336	0.009		
Human-like	Narrator emotion	8.760	0.004	0.081	0.248	0.863
	SLS	0.460	0.499	0.005		

Engaging	Interaction	0.001	0.974	0.000	1.138	0.337
	Narrator emotion	4.562	0.035	0.044		
	SLS	0.013	0.910	0.000		
	Interaction	0.043	0.836	0.000		
Participants' affective state						
Positive activation	Narrator emotion	2.442	0.121	0.024	0.958	0.416
	SLS	0.016	0.898	0.000		
	Interaction	1.370	0.245	0.014		
Negative activation	Narrator emotion	0.151	0.699	0.002	1.066	0.367
	SLS	0.048	0.827	0.000		
	Interaction	7.059	0.009	0.067		
Valence	Narrator emotion	0.109	0.742	0.001	0.657	0.581
	SLS	0.001	0.974	0.000		
	Interaction	0.816	0.368	0.008		
Activation level [†]	Narrator emotion	0.092	0.763	0.001	2.021	0.115
	SLS	0.122	0.727	0.001		
	Interaction	1.088	0.299	0.011		
Valence [†]	Narrator emotion	0.048	0.828	0.000	0.224	0.880
	SLS	0.740	0.392	0.007		
	Interaction	2.461	0.120	0.024		
Interest and motivation						
Situational interest	Narrator emotion	0.001	0.976	0.000	0.201	0.895
	SLS	0.080	0.778	0.001		
	Interaction	2.327	0.130	0.023		
Interest (delayed) [‡]	Narrator emotion	1.604	0.215	0.052	0.966	0.420
	SLS	2.072	0.161	0.067		
	Interaction	0.023	0.880	0.001		
Intrinsic motivation	Narrator emotion	0.777	0.380	0.008	1.294	0.281
	SLS	0.321	0.572	0.003		
	Interaction	0.943	0.334	0.010		
Learners' experience						
Paying attention	Narrator emotion	0.113	0.737	0.001	1.542	0.208
	SLS	1.699	0.195	0.017		
	Interaction	3.015	0.086	0.030		
Difficulty	Narrator emotion	0.315	0.576	0.003	0.638	0.592

Exerting more effort			SLS	0.065	0.800	0.001	0.688	0.561
			Interaction	0.032	0.858	0.000		
			Narrator emotion	0.001	0.973	0.000		
Enjoyment			SLS	0.160	0.690	0.002	1.045	0.376
			Interaction	2.003	0.160	0.020		
			Narrator emotion	0.091	0.764	0.001		
More lessons like this			SLS	0.011	0.918	0.000	0.522	0.668
			Interaction	1.557	0.215	0.016		
			Narrator emotion	0.898	0.346	0.009		
			SLS	0.434	0.511	0.004		
			Interaction	1.656	0.201	0.017		
			Cognitive outcomes					
Intrinsic load	cognitive	Narrator emotion	0.568	0.453	0.006	0.716	0.544	
		SLS	0.342	0.560	0.003			
		Interaction	3.649	0.059	0.036			
Extraneous load	cognitive	Narrator emotion	1.785	0.185	0.018	1.586	0.197	
		SLS	2.771	0.099	0.027			
		Interaction	0.008	0.930	0.000			
Germane load	cognitive	Narrator emotion	0.241	0.625	0.002	0.717	0.544	
		SLS	0.066	0.798	0.001			
		Interaction	9.677	0.002	0.090			
Mental effort average		Narrator emotion	0.000	0.998	0.000	0.361	0.781	
		SLS	0.014	0.907	0.000			
		Interaction	0.660	0.418	0.007			
Immediate testing								
Knowledge		Narrator emotion	4.015	0.048	0.040	0.961	0.414	
		SLS	1.625	0.205	0.016			
		Interaction	1.163	0.284	0.012			
Retention		Narrator emotion	1.674	0.199	0.017	1.400	0.247	
		SLS	1.771	0.186	0.018			
		Interaction	0.612	0.436	0.006			
Transfer		Narrator emotion	5.934	0.017	0.058	0.366	0.778	
		SLS	0.370	0.544	0.004			
		Interaction	1.328	0.252	0.014			
Certainty		Narrator emotion	0.000	0.997	0.000	1.037	0.379	

	SLS	0.414	0.521	0.004		
	Interaction	2.361	0.128	0.024		
Certainty in correct answers	Narrator emotion	0.001	0.981	0.000	0.222	0.881
	SLS	0.070	0.793	0.001		
	Interaction	4.353	0.040	0.043		
Certainty in incorrect answers	Narrator emotion	0.014	0.906	0.000	2.598	0.056
	SLS	0.193	0.662	0.002		
	Interaction	1.753	0.189	0.018		
Self-evaluation	Narrator emotion	0.237	0.627	0.002	1.194	0.316
	SLS	1.231	0.270	0.013		
	Interaction	0.133	0.716	0.001		
Delayed testing [‡]						
Knowledge	Narrator emotion	0.699	0.410	0.024	0.831	0.486
	SLS	0.212	0.649	0.007		
	Interaction	0.383	0.541	0.013		
Retention	Narrator emotion	0.124	0.727	0.004	0.365	0.779
	SLS	0.013	0.909	0.000		
	Interaction	0.000	0.989	0.000		
Transfer	Narrator emotion	2.358	0.136	0.075	0.999	0.405
	SLS	2.003	0.168	0.065		
	Interaction	2.700	0.111	0.085		
Certainty	Narrator emotion	1.947	0.172	0.056	0.719	0.546
	SLS	0.011	0.918	0.000		
	Interaction	0.017	0.897	0.001		
Certainty in correct answers	Narrator emotion	2.932	0.096	0.082	1.025	0.392
	SLS	0.001	0.970	0.000		
	Interaction	0.180	0.674	0.005		
Certainty in incorrect answers	Narrator emotion	1.271	0.268	0.037	0.248	0.862
	SLS	0.028	0.867	0.001		
	Interaction	0.000	0.989	0.000		
Self-evaluation	Narrator emotion	0.907	0.349	0.030	2.660	0.063
	SLS	0.039	0.844	0.001		
	Interaction	0.176	0.678	0.006		

Note. * $df_1 = 1$, $df_2 = 99$; ** $df_1 = 3$, $df_2 = 105$; [†] – * $df_1 = 1$, $df_2 = 100$; ** $df_1 = 3$, $df_2 = 105$; [‡] – * $df_1 = 1$, $df_2 = 33$; ** $df_1 = 3$, $df_2 = 39$; covariates: prior tested knowledge, prior interest, LexTALE, and PANAVA-KS / activation level and valence baseline measures

6.18 Appendix 18: List of study programs for participants in Study 2

University of Primorska:

- Applicative Kinesiology
- Biopsychology
- Computer Science
- Cultural Heritage
- Geography
- Italian Studies
- Language and Interculturality
- Management
- Pedagogy
- Physiotherapy
- Prevention for health
- Primary School Teaching
- Renewable Materials for Healthy Built Environments
- Slovene Studies
- Sustainable Built Environments

Oregon State University:

- Architecture
- Civil Engineering
- Environmental Science
- Forestry
- Natural Resources
- Renewable Materials
- Sustainable Forest Management
- Tourism, Recreation and Adventure Leadership
- Wood Innovation for Sustainability
- Wood Science
- Wood Science and Engineering

Undisclosed:

- Business Administration
- Dental Medicine
- Wood Technology

University in Ljubljana:

- Architecture
- Bioinformatics
- Biology
- Biosciences
- Marketing Communications and Public Relations
- Nursing
- Quantitative Finance and Actuarial Sciences
- Sanitary Engineering
- Social pedagogy
- Special and rehabilitation pedagogy
- Wood Engineering

6.19 Appendix 19: Correlation matrix (Pearson r) of Study 2 outcome variables

	PA	NA	VA	ACT	VAL	INT	IND	MOT	ATT	DIF	EFF	ENJ	LES	ICL	ECL	GCL	ME	SEL	K	R	T	SE	KD	RD	TD	SED	O	C	E	A
PA	—																													
NA	0.35	—																												
VA	0.52	0.49	—																											
ACT	0.37	0.08	0.23	—																										
VAL	0.38	0.34	0.43	0.33	—																									
INT	0.26	0.18	0.28	0.26	0.29	—																								
IND	0.10	0.00	0.13	0.23	0.20	0.72	—																							
MOT	0.27	0.17	0.29	0.31	0.31	0.85	0.66	—																						
ATT	0.27	0.20	0.34	0.27	0.26	0.75	0.45	0.72	—																					
DIF	0.12	0.02	0.05	0.02	0.05	0.23	0.12	0.13	0.15	—																				
EFF	0.00	0.07	0.10	0.07	0.03	0.07	0.00	0.15	0.23	0.43	—																			
ENJ	0.22	0.18	0.28	0.21	0.28	0.81	0.65	0.81	0.73	0.19	0.11	—																		
LES	0.27	0.18	0.25	0.26	0.29	0.76	0.59	0.84	0.68	0.15	0.14	0.77	—																	
ICL	0.09	0.04	0.06	0.03	0.01	0.09	0.07	0.00	0.01	0.40	0.37	0.07	0.04	—																
ECL	0.26	0.14	0.29	0.12	0.19	0.46	0.23	0.46	0.46	0.41	0.15	0.46	0.45	0.41	—															
GCL	0.12	0.12	0.24	0.05	0.17	0.40	0.19	0.38	0.46	0.02	0.25	0.35	0.38	0.12	0.21	—														
ME	0.04	0.05	0.14	0.15	0.16	0.19	0.12	0.16	0.25	0.21	0.56	0.16	0.13	0.26	0.03	0.27	—													
SEL	0.18	0.14	0.23	0.18	0.19	0.53	0.35	0.51	0.49	0.31	0.02	0.50	0.45	0.21	0.47	0.32	0.16	—												
K	0.02	0.05	0.08	0.03	0.14	0.31	0.26	0.26	0.24	0.22	0.11	0.29	0.22	0.18	0.18	0.24	0.00	0.27	—											
R	0.05	0.06	0.11	0.06	0.15	0.29	0.24	0.24	0.23	0.22	0.09	0.29	0.22	0.16	0.18	0.19	0.01	0.27	0.95	—										
T	0.02	0.00	0.02	0.03	0.06	0.24	0.23	0.19	0.20	0.18	0.08	0.21	0.14	0.19	0.14	0.21	0.04	0.21	0.80	0.57	—									
SE	0.04	0.01	0.01	0.05	0.02	0.37	0.39	0.24	0.25	0.30	0.04	0.32	0.19	0.15	0.14	0.21	0.14	0.42	0.36	0.34	0.29	—								
KD	0.10	0.01	0.06	0.26	0.14	0.32	0.30	0.26	0.22	0.23	0.08	0.22	0.19	0.23	0.22	0.20	0.01	0.34	0.85	0.78	0.73	0.33	—							
RD	0.10	0.06	0.07	0.26	0.15	0.31	0.26	0.26	0.24	0.21	0.05	0.23	0.21	0.21	0.24	0.18	0.02	0.33	0.81	0.79	0.62	0.31	0.96	—						
TD	0.07	0.09	0.02	0.19	0.09	0.26	0.29	0.18	0.12	0.21	0.11	0.17	0.11	0.20	0.12	0.20	0.01	0.29	0.71	0.59	0.75	0.30	0.84	0.64	—					
SED	0.13	0.03	0.15	0.16	0.09	0.40	0.36	0.32	0.25	0.24	0.03	0.34	0.29	0.16	0.20	0.23	0.10	0.49	0.39	0.39	0.28	0.77	0.37	0.34	0.34	—				
O	0.06	0.02	0.07	0.03	0.06	0.15	0.06	0.17	0.16	0.23	0.07	0.22	0.10	0.14	0.16	0.03	0.03	0.23	0.17	0.21	0.07	0.15	0.06	0.05	0.05	0.08	—			
C	0.06	0.12	0.03	0.04	0.00	0.17	0.00	0.12	0.19	0.04	0.04	0.11	0.14	0.06	0.13	0.17	0.02	0.15	0.01	0.00	0.03	0.05	0.05	0.05	0.04	0.05	0.25	—		
E	0.01	0.01	0.04	0.13	0.04	0.03	0.03	0.09	0.02	0.03	0.08	0.08	0.12	0.01	0.01	0.04	0.00	0.04	0.19	0.19	0.12	0.01	0.22	0.20	0.19	0.03	0.14	0.15	—	
A	0.07	0.09	0.15	0.05	0.07	0.17	0.04	0.17	0.25	0.08	0.20	0.21	0.17	0.06	0.10	0.21	0.17	0.09	0.07	0.06	0.10	0.07	0.03	0.02	0.13	0.01	0.05	0.17	0.07	—
N	0.08	0.03	0.02	0.01	0.09	0.11	0.14	0.04	0.03	0.10	0.13	0.12	0.07	0.09	0.11	0.03	0.06	0.00	0.06	0.06	0.04	0.07	0.13	0.13	0.09	0.16	0.08	0.01	0.10	0.07

Note. **$p < 0.001$** ; PA – positive activation change score, NA – negative activation change score, VA – valence (PANAVA-KS) change score, ACT – activation level average change score, VAL – valence average change score, INT – situational interest, IND – interest (delayed), MOT – intrinsic motivation, ATT – paying attention, DIF – difficulty, EFF – effort, ENJ – enjoyment, LES – more lessons, ICL – intrinsic cognitive load, ECL – extraneous cognitive load, GCL – germane cognitive load, ME – general mental effort, SEL – self-evaluated learning, K – knowledge, R – retention, T – transfer, SE – self-evaluation, KD – knowledge (delayed), RD – retention (delayed), TD – transfer (delayed), SED – self-evaluation (delayed), O – openness, C – conscientiousness, E – extraversion, A – agreeableness, N – neuroticism

6.20 Appendix 20: Normality and homogeneity test results for Study 2 outcomes

Table 134: Shapiro-Wilk's normality and Levene's homogeneity tests for Study 2 outcome variables before ANOVAs

	Homogeneity test		Normality test	
	<i>F</i>	<i>p</i>	<i>W</i>	<i>p</i>
Video perception*				
Video pleasantness	0.227	0.797	0.954	< .001
Video activation level	3.639	0.027	0.972	< .001
Participants' affective state*				
Positive activation	1.330	0.266	0.992	0.120
Negative activation	0.443	0.643	0.985	0.003
Valence	0.325	0.722	0.974	< .001
Activation level 1	0.204	0.815	0.973	< .001
Activation level 2	2.165	0.117	0.970	< .001
Activation level 3	1.944	0.145	0.974	< .001
Activation level 4	3.839	0.023	0.976	< .001
Activation level 5	1.914	0.149	0.983	0.001
Activation level average	1.996	0.138	0.993	0.181
Valence 1	1.834	0.162	0.969	< .001
Valence 2	0.995	0.371	0.977	< .001
Valence 3	0.572	0.565	0.978	< .001
Valence 4	1.826	0.163	0.967	< .001
Valence 5	0.205	0.815	0.965	< .001
Valence average	2.155	0.118	0.984	0.002
Interest and motivation				
Situational interest*	0.141	0.869	0.990	0.031

Interest (delayed) [§]	0.949	0.390	0.917	< .001
Intrinsic motivation*	0.114	0.892	0.985	0.003
Learners' experience*				
Paying attention	1.529	0.219	0.971	< .001
Difficulty	3.021	0.050	0.912	< .001
Exerting more effort	3.411	0.034	0.953	< .001
Enjoyment	1.854	0.158	0.964	< .001
More lessons like this	2.975	0.053	0.973	< .001
Cognitive outcomes*				
Intrinsic cognitive load	2.192	0.113	0.976	< .001
Extraneous cognitive load	1.246	0.289	0.971	< .001
Germane cognitive load	3.225	0.041	0.974	< .001
Mental effort 1	0.008	0.992	0.981	< .001
Mental effort 2	0.703	0.496	0.974	< .001
Mental effort 3	0.880	0.416	0.975	< .001
Mental effort 4	1.134	0.323	0.980	< .001
Mental effort 5	1.138	0.322	0.977	< .001
Mental effort average	0.697	0.499	0.991	0.067
Learning outcomes				
Self-evaluated learning*	1.219	0.297	0.958	< .001
<i>Immediate testing</i>				
Knowledge [†]	1.083	0.340	0.994	0.324
Retention [†]	2.477	0.086	0.989	0.024
Transfer [†]	0.131	0.878	0.980	< .001
Certainty [‡]	1.062	0.347	0.957	< .001
Certainty in correct answers [‡]	1.768	0.172	0.940	< .001
Certainty in incorrect answers [‡]	0.626	0.535	0.977	< .001

R Certainty [†]	0.803	0.449	0.963	< .001
R Certainty in correct answers [†]	1.676	0.189	0.947	< .001
R Certainty in incorrect answers [†]	0.095	0.910	0.986	0.005
T Certainty [‡]	1.491	0.227	0.947	< .001
T Certainty in correct answers [‡]	2.268	0.105	0.939	< .001
T Certainty in incorrect answers [‡]	1.258	0.286	0.979	< .001
Self-evaluation [‡]	0.941	0.391	0.962	< .001
<i>Delayed testing[§]</i>				
Knowledge	0.495	0.611	0.988	0.370
Retention	1.071	0.346	0.983	0.150
Transfer	0.541	0.584	0.985	0.235
Certainty	0.782	0.460	0.952	< .001
Certainty in correct answers	1.529	0.221	0.934	< .001
Certainty in incorrect answers	1.798	0.170	0.981	0.098
R Certainty	0.448	0.640	0.961	0.002
R Certainty in correct answers	0.816	0.445	0.938	< .001
R Certainty in incorrect answers	1.093	0.339	0.990	0.546
T Certainty	1.882	0.157	0.949	< .001
T Certainty in correct answers	2.938	0.057	0.951	< .001
T Certainty in incorrect answers	1.831	0.165	0.968	0.007
Self-evaluation	0.832	0.438	0.974	0.022
<i>Note.</i> R – retention, T – transfer; * $df_1 = 2$, $df_2 = 304$; [†] $df_1 = 2$, $df_2 = 299$; [‡] $df_1 = 2$, $df_2 = 298$; [§] $df_1 = 2$, $df_2 = 115$				

6.21 Appendix 21: Descriptive statistics by proficiency – Study 2

Table 135: Descriptive statistics for the main outcome variables divided by lower (LexTALE < 69) and higher (LexTALE > 69) English proficiency group – Study 2

	Lower English proficiency group (LexTALE < 69)			Higher English proficiency group (LexTALE > 69)		
	No music (n = 51/18)	Calm (n = 50/18)	Lively (n = 47/21)	No music (n = 49/22)	Calm (n = 50/17)	Lively (n = 52/22)
Video perception						
Perceived video pleasantness						
<i>M</i> (<i>SD</i>)	5.06 (1.27)	5.72 (0.95)	5.23 (1.35)	4.78 (1.52)	4.92 (1.64)	4.46 (1.32)
Min–Max	2–7	3–7	1–7	2–7	1–7	2–7
Skewness	–0.42	–0.89	–1.60	–0.09	–0.79	0.19
Kurtosis	–0.25	1.22	2.93	–1.03	–0.27	–0.89
Perceived video activation level						
<i>M</i> (<i>SD</i>)	4.02 (1.45)	4.42 (1.25)	4.49 (1.37)	3.78 (1.75)	3.84 (1.52)	4.04 (1.37)
Min–Max	1–6	2–7	1–7	1–7	1–7	1–6
Skewness	–0.45	–0.20	–0.38	0.16	0.10	–0.36
Kurtosis	–0.66	–0.58	0	–1.20	–0.58	–0.45
Difference in participants' affective state						
Positive activation change score						
<i>M</i> (<i>SD</i>)	–0.31 (1.18)	–0.16 (1.21)	–0.49 (1.05)	–0.23 (1.09)	–0.35 (1.22)	–0.39 (0.85)
Min–Max	–4–2	–2.50–3	–3–1.75	–3.25–2.25	–3.25– 4.25	–2.75– 1.25
Skewness	–0.53	0.35	0.09	–0.22	1.34	–0.38
Kurtosis	0.98	0.16	–0.30	0.31	4.55	0.42
Negative activation change score						
<i>M</i> (<i>SD</i>)	–0.12 (1.18)	–0.45 (1.05)	–0.05 (1.10)	–0.21 (0.98)	–0.28 (0.97)	–0.13 (0.89)
Min–Max	–6–2	–4.25– 1.25	–3–2.50	–2.50–2.25	–4–1.50	–3–2.50
Skewness	–2.58	–1.11	–0.39	0.44	–1.32	–0.10
Kurtosis	11.85	2.25	0.68	0.81	3.55	2.02

Valence change score						
<i>M (SD)</i>	−0.24 (1.32)	0.01 (1.06)	−0.24 (0.88)	−0.10 (1.19)	−0.07 (1.29)	−0.22 (1.04)
Min–Max	−4–5.50	−2.50–3	−2–2.50	−2.50–4.50	−5–4.50	−3–3
Skewness	1.06	0.58	−0.38	1.08	−0.34	−0.08
Kurtosis	7	1.71	−0.53	3.91	6.43	1.47
Activation level change score						
<i>M (SD)</i>	0.04 (1.58)	0.03 (1.55)	0.04 (0.97)	−0.31 (1.53)	−0.24 (1.75)	0.07 (1.40)
Min–Max	−4.40–4	−3.20– 4.20	−2.40– 2.40	−4–3.20	−5.20–3	−4–3
Skewness	−0.23	0.45	−0.08	−0.31	−0.56	−0.46
Kurtosis	0.88	0.51	0.46	0.27	0.70	0.78
Valence change score						
<i>M (SD)</i>	−0.36 (1.16)	−0.02 (1.53)	−0.06 (1.26)	−0.17 (1.05)	−0.32 (1.71)	−0.31 (1.04)
Min–Max	−5.20–2	−3.20– 3.80	−3–4	−2.80–2.60	−4.20– 4.40	−2.60– 1.80
Skewness	−1.46	0.20	0.41	−0.35	0.69	−0.28
Kurtosis	5.19	−0.32	1.78	0.91	1.32	−0.42
Interest and motivation						
Situational interest						
<i>M (SD)</i>	3.97 (1.21)	4.34 (1.16)	4.36 (1.40)	4.25 (1.28)	4.58 (1.43)	4.09 (1.08)
Min–Max	2–6.33	2.17–6.33	1–7	1–6.67	1–6.33	1.33–6.33
Skewness	0.22	0.05	−0.05	−0.33	−1.21	−0.09
Kurtosis	−0.92	−0.91	−0.29	−0.35	0.52	−0.23
Interest (delayed)						
<i>M (SD)</i>	4.67 (1.53)	3.89 (1.78)	4.71 (1.62)	4.00 (1.66)	5.06 (1.48)	4.27 (1.35)
Min–Max	1–7	1–7	1–7	1–6	1–7	2–7
Skewness	−0.90	−0.02	−0.82	−0.62	−1.43	−0.42
Kurtosis	0.71	−1.29	0.21	−1.10	2.33	−0.13
Intrinsic motivation						
<i>M (SD)</i>	4.08 (1.30)	4.49 (1.14)	4.58 (1.49)	4.26 (1.31)	4.40 (1.45)	4.17 (1.13)

Min–Max	1.88–6.88	2.13–6.88	1.25–7	1.38–6.63	1–6.50	1.38–6.38
Skewness	0.22	–0.25	–0.37	–0.31	–0.71	–0.28
Kurtosis	–0.77	–1.07	–0.53	–0.72	–0.37	–0.41
Learners' experience						
Paying attention						
<i>M (SD)</i>	3.90 (1.63)	4.50 (1.28)	4.34 (1.54)	4.00 (1.58)	4.58 (1.83)	4.15 (1.27)
Min–Max	1–6	2–7	1–7	1–7	1–7	2–6
Skewness	–0.10	0.06	–0.42	0.07	–0.74	–0.06
Kurtosis	–1.23	–0.69	–0.77	–0.93	–0.65	–0.94
Difficulty						
<i>M (SD)</i>	3.31 (1.54)	3.04 (1.24)	2.77 (1.15)	2.39 (1.13)	2.48 (1.39)	2.79 (1.14)
Min–Max	1–7	1–6	1–6	1–5	1–7	1–6
Skewness	0.47	0.32	0.84	0.87	1.21	0.93
Kurtosis	–0.14	–0.67	0.58	0.32	1.45	0.38
Exerting more effort						
<i>M (SD)</i>	3.76 (1.63)	3.86 (1.47)	3.30 (1.40)	2.98 (1.52)	3.20 (1.53)	3.73 (1.17)
Min–Max	1–7	1–7	1–6	1–6	1–7	2–6
Skewness	–0.06	0.09	–0.01	0.67	0.44	–0.21
Kurtosis	–0.80	–0.68	–0.95	–0.50	0	–1.21
Enjoyment						
<i>M (SD)</i>	4.06 (1.54)	4.64 (1.44)	4.68 (1.55)	4.51 (1.56)	4.74 (1.59)	4.40 (1.12)
Min–Max	1–7	2–7	1–7	1–7	1–7	2–7
Skewness	–0.03	–0.27	–0.72	–0.40	–0.95	–0.09
Kurtosis	–1.09	–0.76	–0.08	–0.77	0.23	–0.44
More lessons like this						
<i>M (SD)</i>	3.88 (1.82)	4.74 (1.38)	4.60 (1.66)	4.18 (1.84)	4.46 (1.75)	4.12 (1.32)
Min–Max	1–7	2–7	1–7	1–7	1–7	2–7
Skewness	0.10	–0.33	–0.32	–0.30	–0.77	0.04
Kurtosis	–1.06	–0.59	–0.57	–0.96	–0.54	–0.79
Cognitive outcomes						

Intrinsic cognitive load						
<i>M (SD)</i>	3.86 (1.20)	3.55 (1.43)	3.51 (1.20)	3.14 (1.28)	3.45 (1.37)	3.71 (1.12)
Min–Max	1.50–6	1.50–7	1–6	1–6.50	1–6	1.50–6
Skewness	0.19	0.35	0.25	0.73	–0.03	0
Kurtosis	–0.73	–0.72	–0.51	–0.05	–0.88	–0.66
Extraneous cognitive load						
<i>M (SD)</i>	3.21 (1.27)	2.85 (1.09)	2.85 (1.21)	2.95 (1.39)	2.77 (1.25)	3.19 (1.08)
Min–Max	1–6.67	1–6	1–6	1–6.67	1–6.67	1.33–6.33
Skewness	0.43	0.46	0.67	0.71	0.64	0.49
Kurtosis	0.12	0.03	–0.05	–0.38	0.45	0.15
Germane cognitive load						
<i>M (SD)</i>	4.72 (1.40)	4.88 (1.15)	4.79 (1.33)	4.77 (1.27)	5.08 (1.21)	4.91 (0.88)
Min–Max	1–7	2.50–7	1–7	2.50–7	1–7	3–7
Skewness	–0.66	–0.23	–0.95	0.10	–1.14	–0.11
Kurtosis	–0.09	–0.47	0.88	–0.83	1.72	–0.01
Mental effort (average)						
<i>M (SD)</i>	4.51 (1.62)	4.90 (1.46)	4.35 (1.55)	4.12 (1.33)	4.45 (1.64)	4.56 (1.18)
Min–Max	1.40–8.80	1–8	1–7.80	1.60–7.80	1–7.80	1–7
Skewness	0.16	–0.49	–0.24	0.54	–0.30	–0.53
Kurtosis	0.16	0.51	–0.52	0.24	–0.31	0.66
Immediate testing						
Knowledge						
<i>M (SD)</i>	13.86 (4.12)	13.58 (4.32)	14.57 (4.08)	17.43 (3.86)	18.24 (4.28)	17.71 (5.34)
Min–Max	3–21	4–24	6–24	9–25	7–26	7–28
Skewness	–0.25	0.07	0.13	–0.13	–0.31	–0.12
Kurtosis	–0.31	–0.03	–0.16	–0.42	–0.27	–0.78
Retention						
<i>M (SD)</i>	9.02 (3.02)	8.84 (3.38)	9.21 (3.01)	11.35 (2.72)	11.96 (3.45)	11.69 (3.84)
Min–Max	3–14	2–17	4–18	6–16	3–18	5–18

Skewness	−0.01	0.14	0.57	−0.21	−0.16	0.02
Kurtosis	−0.88	−0.38	0.24	−0.35	−0.42	−1.17
Transfer						
<i>M</i> (<i>SD</i>)	4.84 (1.82)	4.74 (1.47)	5.36 (1.75)	6.08 (1.64)	6.28 (1.51)	6.02 (1.97)
Min–Max	0–8	1–8	1–8	3–10	3–9	2–10
Skewness	−0.53	−0.25	−0.56	0.04	−0.31	−0.09
Kurtosis	0.22	0.04	−0.48	−0.40	−0.60	−0.23
Certainty						
<i>M</i> (<i>SD</i>)	55.13 (21.01)	56.91 (21.60)	64.69 (17.94)	67.26 (13.48)	73.32 (16.25)	70.05 (17.11)
Min–Max	9.14–96.55	0–86.55	14.52–94.83	37.07–88.79	15.86–99.14	6.90–97.41
Skewness	−0.14	−0.88	−0.41	−0.48	−1.36	−1.07
Kurtosis	−0.76	0.23	0.01	−0.52	2.51	2.48
Certainty in correct answers						
<i>M</i> (<i>SD</i>)	55.99 (21.87)	59.82 (23.54)	66.85 (18.97)	71.43 (13.45)	77.92 (14.90)	73.69 (17.40)
Min–Max	9–99.76	0–95.38	14.62–94.71	40–92	27.69–100	6.25–97.22
Skewness	−0.05	−0.83	−0.70	−0.38	−1.34	−1.52
Kurtosis	−0.86	0.07	0.06	−0.60	2.18	3.75
Certainty in incorrect answers						
<i>M</i> (<i>SD</i>)	53.78 (20.75)	53.56 (20.30)	61.99 (17.40)	59.68 (14.90)	66.40 (17.17)	63.86 (17.06)
Min–Max	9.29–88.13	0–84.74	14.44–96.15	15.91–86.11	6.25–100	7.69–100
Skewness	−0.23	−0.82	−0.22	−0.66	−0.98	−0.43
Kurtosis	−0.82	0.17	−0.08	0.38	2.13	1.14
R Certainty						
<i>M</i> (<i>SD</i>)	54.57 (22.01)	56.05 (20.98)	63.90 (17.72)	67.14 (13.78)	73.50 (15.89)	69.14 (17.91)
Min–Max	9.47–94.74	0–88.37	14.05–93.44–2	40–90.79	21.58–98.68	5.26–96.05
Skewness	−0.13	−0.84	−0.33	−0.30	−1.26	−0.86
Kurtosis	−1.03	0.20	0.10	−0.79	2	1.81

R Certainty in correct answers						
<i>M (SD)</i>	55.50 (22.01)	58.89 (20.98)	66.97 (17.72)	71.41 (14.44)	78.40 (15.70)	73.77 (17.47)
Min–Max	8.89–99.62	0–93.33	15.70– 94.44	40–96.67	34.44– 100	10–97.33
Skewness	–0.10	–0.81	–0.72	–0.29	–1.31	–1.31
Kurtosis	–0.82	0.14	–0.02	–0.62	1.31	2.96
R Certainty in incorrect answers						
<i>M (SD)</i>	52.83 (23.14)	52.40 (22.61)	60.95 (19.14)	59.02 (15.89)	64.81 (16.91)	61.05 (19.22)
Min–Max	10–90	0–83.85	12.22– 100	18.75– 89.29	10–100	0–100
Skewness	–0.17	–0.65	–0.08	–0.40	–0.71	–0.35
Kurtosis	–1.06	0.06	0.08	–0.01	1.43	0.68
T Certainty						
<i>M (SD)</i>	56.21 (23.03)	58.57 (24.60)	66.19 (20.77)	67.48 (15.83)	72.96 (18.91)	71.78 (17.33)
Min–Max	2.80–100	0–100	15.40– 100	27.50– 91.30	5–100	10–100
Skewness	–0.30	–0.70	–0.57	–0.79	–1.17	–1.40
Kurtosis	–0.60	–0.11	–0.39	0.18	2.16	2.81
T Certainty in correct answers						
<i>M (SD)</i>	57.80 (25.38)	60.87 (27.72)	68.82 (21.94)	71.21 (16.95)	76.22 (18.63)	73.97 (19.13)
Min–Max	4–100	0–100	11–100	35.71– 96.43	12.50– 100	0–100
Skewness	–0.18	–0.54	–0.84	–0.62	–1.20	–1.49
Kurtosis	–0.93	–0.54	0.11	–0.53	1.75	3.40
T Certainty in incorrect answers						
<i>M (SD)</i>	55.02 (23.58)	56.07 (24.26)	63.21 (22.66)	60.96 (17.85)	68.93 (21.22)	67.85 (16.97)
Min–Max	2.29–100	0–100	17.29– 100	8.33–87.80	0–100	25–100
Skewness	–0.08	–0.42	–0.22	–0.78	–0.70	–0.74
Kurtosis	–0.41	–0.21	–0.88	0.51	0.80	0.57
Self–evaluated learning						

<i>M (SD)</i>	3.78 (1.22)	4.32 (0.96)	4.30 (1.12)	4.39 (1.38)	4.68 (1.20)	4.33 (1.06)
Min–Max	1–6	2–7	2–7	1–7	1–7	1–6
Skewness	–0.53	0.18	0.24	–0.20	–0.59	–0.19
Kurtosis	0.62	1.11	0.57	–0.13	2.17	0.64
Self–evaluated test performance						
<i>M (SD)</i>	3.59 (1.25)	3.52 (1.25)	3.79 (1.35)	3.61 (1.10)	4.22 (1.37)	3.98 (1.04)
Min–Max	1–7	1–7	1–7	1–6	1–7	1–6
Skewness	0.01	0.08	0.46	–0.55	–0.17	–0.73
Kurtosis	0.30	0.39	0.17	0.48	0.08	1.67
Delayed testing						
Knowledge						
<i>M (SD)</i>	14 (3.84)	13.61 (4.58)	14.29 (4.11)	16.68 (4.59)	19 (4.40)	17.73 (4.57)
Min–Max	6–21	7–24	6–23	9–26	6–24	11–25
Skewness	–0.14	0.37	0.16	–0.11	–1.54	0.05
Kurtosis	0.16	0.17	0.18	–0.56	3.86	–1.26
Retention						
<i>M (SD)</i>	9.35 (2.85)	8.67 (3.36)	9.14 (3.09)	10.59 (3.20)	12.24 (3.36)	11.50 (3.26)
Min–Max	4–15	3–16	4–17	4–16	3–17	6–17
Skewness	–0.20	0.45	0.64	–0.50	–1.16	–0.04
Kurtosis	0.02	0.62	0.67	–0.24	2.30	–1.29
Transfer						
<i>M (SD)</i>	4.65 (1.54)	4.94 (1.86)	5.14 (1.46)	6.09 (1.66)	6.76 (1.60)	6.23 (1.77)
Min–Max	2–7	1–9	2–8	4–10	3–9	2–9
Skewness	–0.26	0.27	0.15	0.46	–0.81	–0.15
Kurtosis	–0.65	0.93	0.56	–0.23	0.77	0.16
Certainty						
<i>M (SD)</i>	54.86 (19.23)	55.80 (21.62)	61.31 (18.76)	60.29 (18.45)	74.28 (16.82)	69.17 (15.89)
Min–Max	17.07–85.69	25.62–96.55	22.76–92.10	5.86–83.28	25.69–99.14	23.28–91.38
Skewness	–0.55	0.03	–0.30	–1.20	–1.48	–1.12

Kurtosis	−0.60	−1.01	−0.54	2.34	3.57	1.93
Certainty in correct answers						
<i>M (SD)</i>	57.14 (20.99)	56.65 (22.72)	65.53 (19.19)	63.12 (20.46)	77.79 (16.94)	73.73 (15.52)
Min–Max	19.55– 87.50	27.86– 93.33	24.55– 91.67	3.89–89.17	27.50– 100	26.92–96
Skewness	−0.42	−0.10	−0.67	−1.14	−1.80	−1.21
Kurtosis	−1.10	−1.54	−0.25	1.90	4.23	2.67
Certainty in incorrect answers						
<i>M (SD)</i>	51.89 (18.27)	53.32 (21.61)	56.45 (17.89)	54.72 (15.71)	66.92 (18.34)	62.09 (15.25)
Min–Max	15.56– 81.67	24.91– 100	21.67– 92.57	6.75–74.17	25.22– 97.92	20.31– 81.67
Skewness	−0.55	0.39	0.22	−1.39	−0.48	−0.98
Kurtosis	−0.56	−0.43	−0.11	2.96	0.40	1.10
R Certainty						
<i>M (SD)</i>	53.01 (18.72)	54.59 (20.85)	58.81 (19.23)	58.11 (18.58)	73.26 (16.70)	67.31 (17.23)
Min–Max	16.05– 80.26	25–94.74	20–93.21	7.89–85.53	25–98.68	21.05– 92.11
Skewness	−0.46	0.12	−0.24	−0.92	−1.47	−0.95
Kurtosis	−0.76	−0.93	−0.44	1.14	3.58	1.16
R Certainty in correct answers						
<i>M (SD)</i>	55.99 (20.34)	54.42 (21.84)	62.03 (21.40)	61.98 (20.27)	78.01 (17.11)	73.04 (17.74)
Min–Max	18.75– 83.08	25–90	20–97.73	8.75–89.06	25–100	25–96.88
Skewness	−0.39	−0.03	−0.41	−0.90	−2.08	−0.98
Kurtosis	−1.18	−1.55	−0.65	0.68	5.49	1.24
R Certainty in incorrect answers						
<i>M (SD)</i>	49.05 (17.88)	52.06 (21.39)	54.63 (18.54)	51.48 (16.02)	63.59 (18.37)	57.84 (15.88)
Min–Max	14.09–75	25–100	20–95	7.67–71.25	25–96.88	18.75–79
Skewness	−0.31	0.58	0.37	−0.98	−0.13	−0.91
Kurtosis	−0.74	−0.15	0.07	1.14	0.02	0.36
T Certainty						

<i>M (SD)</i>	58.37 (22.39)	58.09 (24.34)	66.05 (19.32)	64.43 (19.81)	76.24 (18.11)	72.72 (14.64)
Min–Max	19–96	25–100	28–93	2–92	27–100	27.50– 94.90
Skewness	–0.30	–0.02	–0.41	–1.44	–1.23	–1.29
Kurtosis	–0.74	–1.29	–0.73	3.64	2.18	3.20
T Certainty in correct answers						
<i>M (SD)</i>	60.16 (24.37)	59.63 (26.55)	71.48 (21.15)	65.90 (22.43)	77.14 (17.49)	75.32 (13.90)
Min–Max	21.67– 95.71	25–100	32.50– 100	0–100	30–100	29.17–95
Skewness	–0.14	–0.09	–0.55	–1.02	–1.22	–1.57
Kurtosis	–1.21	–1.57	–0.92	2.30	2.07	5.03
T Certainty in incorrect answers						
<i>M (SD)</i>	56.28 (22.03)	55.80 (24.12)	60.73 (20.34)	59.59 (18.50)	74.84 (21.25)	69.78 (18.06)
Min–Max	17.86– 96.67	24.67– 100	25–100	4–82.50	25.71– 100	25–94.88
Skewness	–0.18	0.09	0.15	–1.39	–0.80	–0.84
Kurtosis	–0.63	–1.16	–0.73	2.81	0.02	0.39
Self-evaluated test performance						
<i>M (SD)</i>	3.28 (1.18)	3.50 (1.20)	3.43 (1.33)	3.36 (1.22)	4.47 (1.37)	4.09 (0.75)
Min–Max	1–5	1–6	1–7	1–6	1–7	3–6
Skewness	–0.37	–0.23	0.95	–0.09	–0.50	0.59
Kurtosis	–1	0.39	2.06	–0.32	1.64	0.86

6.22 Appendix 22: ANCOVA comparisons for low proficiency group – Study 2

Table 136: ANCOVA comparisons on all main dependable variables on the lower English proficiency group (LexTALE < 69) – Study 2

	ANCOVA*			Homogeneity test**		Normality test	
	<i>F</i>	<i>p</i>	η^2p	<i>F</i>	<i>p</i>	<i>W</i>	<i>p</i>
Video perception							
Pleasantness	4.60	0.012	0.06	1.11	0.332	0.97	< .001
Activation level	1.45	0.237	0.02	0.62	0.538	0.97	0.002
Differences in affective state							
Positive activation	1.13	0.328	0.02	1.32	0.272	0.99	0.513
Negative activation	4.27	0.016	0.06	1.23	0.295	0.98	0.015
Valence	2.76	0.067	0.04	2.18	0.117	0.96	< .001
Activation level	1.30	0.275	0.02	1.92	0.151	0.99	0.159
Valence ^Q	5.10	0.007		3.80	0.025	0.99	0.124
Interest and motivation							
Situational interest	2.03	0.135	0.03	0.17	0.847	0.99	0.865
Interest (delayed)	0.06	0.943	0.00	1.72	0.189	0.98	0.533
Intrinsic motivation	1.75	0.179	0.03	0.01	0.987	0.99	0.437
Learners' experience							
Paying attention	3.17	0.045	0.05	0.01	0.991	0.98	0.037
Difficulty ^Q	2.00	0.139		3.50	0.033	0.98	0.032
Exerting more effort	2.88	0.060	0.04	0.35	0.706	0.99	0.417
Enjoyment	2.78	0.066	0.04	1.35	0.263	0.99	0.345
More lessons like this	5.12	0.007	0.07	2.30	0.104	0.99	0.131
Cognitive outcomes							
Intrinsic cognitive load	1.23	0.296	0.02	0.56	0.571	0.98	0.021
Extraneous cognitive load	1.68	0.191	0.02	0.71	0.493	0.97	0.001
Germane cognitive load	0.77	0.466	0.01	0.62	0.541	0.97	0.002

Mental effort (average)	3.81	0.025	0.05	0.23	0.795	0.99	0.368
Immediate testing							
Knowledge	0.72	0.489	0.01	1.20	0.304	0.99	0.170
Retention	0.21	0.809	0.00	0.72	0.488	0.99	0.769
Transfer	1.31	0.273	0.02	1.60	0.205	0.98	0.018
Certainty	1.91	0.152	0.03	1.53	0.221	0.99	0.337
Certainty in correct answers	2.24	0.111	0.03	1.14	0.324	0.99	0.282
Certainty in incorrect answers	1.61	0.204	0.02	0.79	0.454	0.99	0.522
R Certainty	1.81	0.167	0.03	0.56	0.572	0.99	0.397
R Certainty in correct answers	2.61	0.077	0.04	0.42	0.657	0.99	0.304
R Certainty in incorrect answers	1.60	0.206	0.02	0.32	0.724	0.99	0.509
T Certainty	1.52	0.223	0.02	1.27	0.285	0.98	0.020
T Certainty in correct answers	1.71	0.185	0.03	2.25	0.109	0.98	0.046
T Certainty in incorrect answers	0.83	0.439	0.01	0.68	0.507	0.99	0.145
Self-evaluated learning	4.50	0.013	0.06	2.23	0.111	0.98	0.051
Self-evaluated test performance	0.03	0.972	0.00	0.08	0.924	0.99	0.746
Delayed testing [†]							
Knowledge	0.19	0.828	0.01	0.28	0.760	0.96	0.087
Retention	0.22	0.802	0.01	0.32	0.724	0.98	0.363
Transfer	0.05	0.949	0.00	0.58	0.566	0.96	0.039
Certainty	0.37	0.691	0.02	0.14	0.867	0.98	0.566
Certainty in correct answers	0.48	0.622	0.02	0.10	0.904	0.98	0.301
Certainty in incorrect answers	0.32	0.731	0.01	0.41	0.669	0.99	0.733
R Certainty	0.29	0.753	0.01	0.15	0.862	0.97	0.273
R Certainty in correct answers	0.22	0.802	0.01	0.23	0.796	0.97	0.150

R Certainty in incorrect answers	0.35	0.703	0.02	0.57	0.571	0.98	0.439
T Certainty	0.47	0.629	0.02	0.22	0.806	0.98	0.698
T Certainty in correct answers	0.89	0.419	0.04	0.09	0.917	0.98	0.579
T Certainty in incorrect answers	0.32	0.729	0.01	0.32	0.727	0.98	0.558
Self-evaluated test performance	0.93	0.403	0.04	0.09	0.916	0.96	0.040

Note. * $df_1 = 2$, $df_2 = 134$; ** $df_1 = 2$, $df_2 = 145$; † * $df_1 = 2$, $df_2 = 43$; ** $df_1 = 2$, $df_2 = 54$; R – retention, T – transfer, ^Q – Quade test results reported instead of ANCOVA due to the homogeneity of variances assumption not being met ($df = 145$); covariates included are prior tested knowledge, prior interest, LexTALE score, the five personality characteristics, and PANAVA–KS baseline measures, except for the Activation level and Valence variables, which had the activation level and valence baseline measures

6.23 Appendix 23: ANCOVA comparisons for high proficiency group – Study 2

Table 137: ANCOVA comparisons on all main dependable variables in the higher English proficiency group (LexTALE > 69) – Study 2

	ANCOVA*			Homogeneity test**		Normality test	
	<i>F</i>	<i>p</i>	η^2p	<i>F</i>	<i>p</i>	<i>W</i>	<i>p</i>
Video perception							
Pleasantness	0.63	0.536	0.01	0.78	0.462	0.99	0.419
Activation level	0.81	0.447	0.01	2.04	0.134	0.99	0.196
Differences in affective state							
Positive activation	0.49	0.616	0.01	2.02	0.136	0.99	0.192
Negative activation	0.43	0.650	0.01	0.14	0.871	0.99	0.578
Valence	0.14	0.865	0.00	0.06	0.943	0.98	0.028
Activation level ^Q	0.33	0.723		3.22	0.043	0.98	0.019
Valence	0.39	0.676	0.01	2.02	0.137	0.98	0.020
Interest and motivation							
Situational interest	0.53	0.591	0.01	0.96	0.385	0.99	0.407
Interest (delayed)	2.61	0.084	0.10	2.26	0.114	0.98	0.336
Intrinsic motivation	0.00	0.999	0.00	2.16	0.118	0.98	0.020
Learners' experience							
Paying attention ^Q	2.07	0.130		3.15	0.046	0.99	0.266
Difficulty	1.18	0.311	0.02	0.19	0.825	0.96	< .001
Exerting more effort	4.11	0.019	0.06	1.20	0.306	0.99	0.465
Enjoyment	0.06	0.945	0.00	1.34	0.264	0.98	0.059
More lessons like this	0.04	0.957	0.00	1.19	0.308	0.99	0.423
Cognitive outcomes							
Intrinsic cognitive load	1.90	0.154	0.03	1.05	0.353	0.99	0.805
Extraneous cognitive load	0.39	0.675	0.01	1.79	0.170	0.96	< .001
Germane cognitive load	0.58	0.564	0.01	2.28	0.106	0.98	0.016

Mental effort (average) ^Q	3.05	0.050		3.72	0.026	0.99	0.591
Immediate testing							
Knowledge	0.05	0.953	0.00	0.80	0.451	0.99	0.641
Retention	0.17	0.840	0.00	1.37	0.258	0.99	0.684
Transfer	0.14	0.874	0.00	0.88	0.417	0.99	0.204
Certainty	0.73	0.395	0.01	1.56	0.214	0.97	0.001
Certainty in correct answers	1.00	0.370	0.01	1.22	0.297	0.95	< .001
Certainty in incorrect answers	1.08	0.341	0.02	0.87	0.420	0.98	0.044
R Certainty	0.67	0.511	0.01	2.40	0.094	0.98	0.018
R Certainty in correct answers	1.11	0.332	0.02	0.99	0.375	0.97	< .001
R Certainty in incorrect answers	0.53	0.589	0.01	0.66	0.520	0.99	0.169
T Certainty	0.59	0.553	0.01	0.44	0.644	0.95	< .001
T Certainty in correct answers	0.38	0.688	0.01	0.22	0.802	0.96	< .001
T Certainty in incorrect answers	1.55	0.217	0.02	1.78	0.173	0.98	0.036
Self-evaluated learning	0.46	0.634	0.01	1.13	0.325	0.98	0.040
Self-evaluated test performance	2.68	0.072	0.04	2.19	0.116	0.99	0.380
Delayed testing [†]							
Knowledge	0.77	0.468	0.03	2.05	0.138	0.98	0.250
Retention	0.89	0.416	0.04	1.38	0.260	0.97	0.224
Transfer	0.26	0.771	0.01	0.50	0.608	0.99	0.768
Certainty	2.26	0.116	0.09	0.31	0.734	0.94	0.004
Certainty in correct answers	2.66	0.081	0.10	0.71	0.498	0.94	0.004
Certainty in incorrect answers	2.04	0.141	0.08	0.11	0.898	0.97	0.216
R Certainty	2.49	0.094	0.10	0.73	0.488	0.95	0.011
R Certainty in correct answers	2.67	0.080	0.10	1.08	0.348	0.95	0.013

R Certainty in incorrect answers	1.76	0.183	0.07	0.08	0.919	0.99	0.758
T Certainty	1.58	0.216	0.06	0.01	0.994	0.93	0.003
T Certainty in correct answers	1.81	0.174	0.07	0.30	0.739	0.94	0.004
T Certainty in incorrect answers	2.54	0.090	0.10	0.18	0.834	0.95	0.024
Self-evaluated test performance ^Q	3.12	0.051		3.70	0.031	0.98	0.292

Note. * $df_1 = 2$, $df_2 = 137$; ** $df_1 = 2$, $df_2 = 148$; [†] * $df_1 = 2$, $df_2 = 47$; ** $df_1 = 2$, $df_2 = 58$; R – retention, T – transfer, ^Q – Quade test results reported instead of ANCOVA due to the homogeneity of variances assumption not being met ($df = 148$); covariates included are prior tested knowledge, prior interest, LexTALE score, the five personality characteristics, and PANAVA–KS baseline measures, except for the Activation level and Valence variables, which had the activation level and valence baseline measures

6.24 Appendix 24: Descriptive statistics by wood science familiarity – Study 2

Table 138: Descriptive statistics for the main outcome variables divided by lower study program familiarity with wood science – Study 2

	Wood science related programs			Non-wood science related programs		
	No music (<i>n</i> = 33/10)	Calm (<i>n</i> = 36/9)	Lively (<i>n</i> = 37/12)	No music (<i>n</i> = 68/30)	Calm (<i>n</i> = 64/26)	Lively (<i>n</i> = 62/31)
Video perception						
Perceived video pleasantness						
<i>M</i> (<i>SD</i>)	5.12 (1.41)	4.97 (1.70)	5 (1.45)	4.79 (1.40)	5.52 (1.15)	4.73 (1.34)
Min–Max	2–7	1–7	1–7	2–7	2–7	1–7
Skewness	–0.51	–0.99	–0.86	–0.13	–0.93	–0.44
Kurtosis	–0.25	0	0.56	–0.90	0.74	–0.46
Perceived video activation level						
<i>M</i> (<i>SD</i>)	4.36 (1.60)	4.03 (1.50)	4.51 (1.33)	3.65 (1.56)	4.19 (1.37)	4.10 (1.40)
Min–Max	1–6	1–7	1–7	1–7	1–7	1–7
Skewness	–0.64	–0.10	–0.30	0.15	–0.12	–0.36
Kurtosis	–0.71	–0.58	0.25	–0.85	–0.56	–0.53
Difference in participants' affective state						
Positive activation change score						
<i>M</i> (<i>SD</i>)	–0.02 (1.03)	–0.10 (1.23)	–0.22 (0.91)	–0.43 (1.19)	–0.34 (1.20)	–0.58 (0.96)
Min–Max	–2–2.25	–3.25– 4.25	–2–1.25	–4–2	–2.50– 3.25	–3–1.75
Skewness	0.21	0.84	–0.15	–0.54	0.86	–0.07
Kurtosis	0	4.03	–0.68	0.44	1.21	0.35
Negative activation change score						
<i>M</i> (<i>SD</i>)	–0.05 (0.85)	–0.35 (1.01)	–0.09 (0.85)	–0.21 (1.17)	–0.38 (1.01)	–0.09 (1.07)
Min–Max	–2.50–1.75	–4–1	–2–1.75	–6–2.25	–4.25– 1.50	–3–2.50
Skewness	–0.35	–1.28	–0.21	–1.59	–1.17	–0.27
Kurtosis	1.05	3.53	0.01	8.07	2.45	1.25

Valence change score						
<i>M (SD)</i>	0 (1.13)	−0.11 (1.33)	−0.22 (1.05)	−0.29 (1.33)	0.02 (1.08)	−0.24 (0.91)
Min–Max	−2.50–4.50	−5–4.50	−3–1.50	−4–5.50	−3–3	−2–3
Skewness	1.77	−0.16	−1	0.85	0.16	0.54
Kurtosis	7.93	8.02	0.41	4.76	1.58	1.46
Activation level change score						
<i>M (SD)</i>	0.15 (1.44)	0.02 (1.88)	0.39 (1.02)	−0.32 (1.65)	−0.17 (1.52)	−0.15 (1.28)
Min–Max	−3.60–2.80	−5.20– 4.20	−2.40–3	−4.40–4	−3.80– 3.60	−4–2.60
Skewness	−0.16	−0.50	0.06	−0.27	0.11	−0.36
Kurtosis	0.20	1.09	1.08	0.48	0.43	0.86
Valence change score						
<i>M (SD)</i>	0.02 (0.69)	−0.42 (1.41)	0.12 (0.91)	−0.45 (1.28)	−0.03 (1.73)	−0.38 (1.25)
Min–Max	−1.40–1.40	−4.20– 2.60	−2.40–2	−5.20–2.60	−3.40– 4.40	−3–4
Skewness	−0.26	−0.17	−0.54	−0.81	0.54	0.60
Kurtosis	−0.27	0.52	1.18	2.20	0.23	1.61
Interest and motivation						
Situational interest						
<i>M (SD)</i>	4.62 (1.22)	4.81 (1.14)	4.86 (1.29)	3.85 (1.17)	4.27 (1.36)	3.84 (1.05)
Min–Max	2–6.50	1.50–6.33	1–7	1–6.67	1–6.33	1.33–6.67
Skewness	−0.44	−1.21	−0.70	0.08	−0.48	0.15
Kurtosis	−0.63	1.55	1.03	−0.48	−0.49	0.10
Interest (delayed)						
<i>M (SD)</i>	5.50 (0.85)	5.89 (0.60)	5.25 (1.60)	3.90 (1.63)	3.96 (1.71)	4.19 (1.35)
Min–Max	4–7	5–7	1–7	1–6	1–7	2–7
Skewness	0	−0.02	−1.76	−0.50	−0.20	−0.38
Kurtosis	0.11	1.13	4.30	−1.07	−1.05	−0.50
Intrinsic motivation						
<i>M (SD)</i>	4.58 (1.45)	4.53 (1.26)	4.80 (1.40)	3.95 (1.19)	4.40 (1.32)	4.11 (1.21)

Min–Max	1.88–6.88	1–6.50	1.25–7	1.38–6.63	1–6.38	1.88–7
Skewness	–0.48	–0.74	–0.83	0.10	–0.52	0.08
Kurtosis	–0.86	0.32	0.53	–0.57	–0.65	–0.38
Learners' experience						
Paying attention						
<i>M (SD)</i>	4.24 (1.75)	4.58 (1.68)	4.68 (1.43)	3.78 (1.51)	4.52 (1.52)	3.98 (1.32)
Min–Max	1–6	1–7	2–7	1–7	1–7	1–6
Skewness	–0.47	–0.78	–0.46	0.23	–0.35	–0.23
Kurtosis	–1.30	–0.21	–0.71	–0.72	–0.58	–0.84
Difficulty						
<i>M (SD)</i>	2.76 (1.41)	2.39 (1.38)	2.49 (0.96)	2.94 (1.45)	2.97 (1.28)	2.95 (1.21)
Min–Max	1–7	1–7	1–5	1–7	1–6	1–6
Skewness	0.88	1.73	1.23	0.67	0.20	0.67
Kurtosis	0.95	3.41	1.05	–0.11	–0.85	0.17
Exerting more effort						
<i>M (SD)</i>	3.06 (1.58)	3.17 (1.38)	3.43 (1.26)	3.56 (1.62)	3.73 (1.58)	3.58 (1.33)
Min–Max	1–6	1–7	1–6	1–7	1–7	1–6
Skewness	0.55	0.51	0.07	0.13	0.05	–0.31
Kurtosis	–0.68	0.09	–1.05	–0.92	–0.53	–0.88
Enjoyment						
<i>M (SD)</i>	4.61 (1.68)	5 (1.24)	4.97 (1.40)	4.12 (1.47)	4.52 (1.62)	4.27 (1.24)
Min–Max	1–7	2–7	1–7	1–7	1–7	1–7
Skewness	–0.68	–0.38	–0.97	0.02	–0.59	–0.28
Kurtosis	–0.88	–0.45	0.89	–0.76	–0.54	–0.20
More lessons like this						
<i>M (SD)</i>	4.61 (1.85)	4.83 (1.52)	4.76 (1.59)	3.74 (1.75)	4.47 (1.60)	4.10 (1.41)
Min–Max	1–7	1–7	1–7	1–7	1–7	1–7
Skewness	–0.51	–1.20	–0.32	0.08	–0.45	–0.03
Kurtosis	–0.84	0.75	–0.51	–0.96	–0.50	–0.65
Cognitive outcomes						

Intrinsic cognitive load						
<i>M (SD)</i>	3.33 (1.16)	3.28 (1.27)	3.58 (1.10)	3.59 (1.33)	3.63 (1.45)	3.64 (1.20)
Min–Max	2–6	1–6	2–6	1–6.50	1–7	1–6
Skewness	0.64	0.29	0.32	0.27	0.07	0.01
Kurtosis	–0.24	–0.55	–0.47	–0.72	–0.86	–0.68
Extraneous cognitive load						
<i>M (SD)</i>	2.86 (1.33)	2.71 (1.10)	3.06 (1.30)	3.17 (1.33)	2.86 (1.21)	3.01 (1.07)
Min–Max	1–6	1–6.67	1–6.33	1–6.67	1–6	1–5.33
Skewness	0.76	1.27	0.67	0.50	0.24	0.34
Kurtosis	–0.27	3.38	0.27	–0.14	–0.66	–0.70
Germane cognitive load						
<i>M (SD)</i>	4.82 (1.12)	4.99 (1.19)	4.96 (1.18)	4.71 (1.42)	4.98 (1.18)	4.79 (1.07)
Min–Max	2–7	1–7	1–7	1–7	2–7	1.50–7
Skewness	–0.27	–1.14	–1.12	–0.35	–0.45	–0.71
Kurtosis	0.33	2.24	2.68	–0.57	–0.34	0.75
Mental effort (average)						
<i>M (SD)</i>	4.16 (1.27)	4.39 (1.43)	4.43 (1.59)	4.39 (1.58)	4.83 (1.63)	4.48 (1.22)
Min–Max	1.40–7	1.20–8	1–7.80	1.60–8.80	1–7.80	1.80–7
Skewness	–0.27	–0.09	–0.55	0.47	–0.62	–0.17
Kurtosis	0.06	0.33	0.16	0.06	0.11	–0.74
Immediate testing						
Knowledge						
<i>M (SD)</i>	17.09 (4.10)	17.89 (4.24)	18.16 (5.12)	14.93 (4.31)	14.80 (4.89)	15.06 (4.62)
Min–Max	8–24	8–25	10–28	3–25	4–26	6–25
Skewness	–0.20	–0.10	0.04	–0.23	0.05	0.14
Kurtosis	–0.27	–0.64	–1.05	–0.17	–0.36	–0.33
Retention						
<i>M (SD)</i>	11.36 (2.77)	11.69 (3.49)	11.70 (3.74)	9.59 (3.07)	9.67 (3.70)	9.81 (3.47)
Min–Max	6–16	5–17	6–18	3–16	2–18	4–18

Skewness	−0.06	−0.15	0.16	−0.16	0.15	0.49
Kurtosis	−0.67	−0.80	−1.16	−0.76	−0.27	−0.45
Transfer						
<i>M (SD)</i>	5.73 (1.86)	6.19 (1.60)	6.46 (1.74)	5.34 (1.82)	5.13 (1.60)	5.26 (1.84)
Min–Max	1–9	3–9	3–10	0–10	1–9	1–9
Skewness	−0.48	−0.29	0.07	−0.31	−0.16	−0.33
Kurtosis	0.26	−1.08	−0.14	0.52	0.12	−0.58
Certainty						
<i>M (SD)</i>	65.15 (20.28)	73.34 (14.65)	76.93 (16.24)	59.54 (17.90)	60.49 (22.25)	61.88 (16.04)
Min–Max	22.45–96.55	38.62–99.14	6.90–97.41	9.14–90.69	0–88.79	14.52–94.83
Skewness	−0.67	−0.48	−2.28	−0.58	−0.97	−0.28
Kurtosis	−0.67	−0.08	8.61	0.11	0.20	0.29
Certainty in correct answers						
<i>M (SD)</i>	67.71 (21.39)	77.28 (14.76)	79.38 (16.17)	61.94 (18.89)	64.14 (23.44)	65.11 (17.66)
Min–Max	22.25–99.76	38.57–100	6.25–97.22	9–92	0–97.06	14.62–93.75
Skewness	−0.66	−0.91	−2.63	−0.59	−0.99	−0.67
Kurtosis	−0.70	0.64	10.87	−0.03	0.33	0.11
Certainty in incorrect answers						
<i>M (SD)</i>	59.61 (18.99)	67.33 (14.39)	71.89 (16.20)	55.75 (18.26)	55.84 (21.25)	57.65 (15.53)
Min–Max	22.69–88.13	38.67–100	7.69–100	9.29–90.91	0–87.50	14.44–96.15
Skewness	−0.49	0.10	−1.62	−0.48	−0.84	0.19
Kurtosis	−0.88	−0.02	5.70	−0.01	0.08	0.36
R Certainty						
<i>M (SD)</i>	65.09 (21.12)	72.68 (14.84)	76.13 (17.07)	59.07 (18.60)	60.33 (21.94)	61 (16.04)
Min–Max	22.21–94.74	37.89–98.68	5.26–96.05	9.47–92.11	0–88.42	14.05–93.68
Skewness	−0.77	−0.32	−2.04	−0.47	−0.93	−0.11
Kurtosis	−0.56	−0.33	7.12	−0.19	0.09	0.34

R Certainty in correct answers						
<i>M (SD)</i>	67.03 (23.13)	77 (15.59)	78.31 (16.47)	61.93 (19.82)	63.95 (23.31)	65.91 (18.21)
Min–Max	9.20–99.62	37.14– 100	10–97.33	8.89–96.67	0–95.83	15.70– 96.15
Skewness	–0.78	–0.79	–2.04	–0.48	–0.92	–0.68
Kurtosis	–0.05	0.24	7.05	–0.30	0.13	0.14
R Certainty in incorrect answers						
<i>M (SD)</i>	59.79 (19.27)	66.07 (15.10)	71.08 (18.84)	54.46 (18.94)	54.41 (20.54)	54.99 (15.49)
Min–Max	22.50–90	38.33– 100	0–100	10–90	0–90	12.22– 94.44
Skewness	–0.46	0.21	–1.51	–0.32	–0.75	0.27
Kurtosis	–0.65	–0.29	4.47	–0.55	–0.01	0.43
T Certainty						
<i>M (SD)</i>	65.26 (23.22)	74.61 (17.20)	78.46 (16.01)	60.41 (19.18)	60.79 (24.44)	63.56 (18.80)
Min–Max	22–100	37.50– 100	10–100	2.80–91	0–100	15.40– 100
Skewness	–0.66	–0.68	–2.29	–0.81	–0.83	–0.60
Kurtosis	–0.84	–0.24	8.49	0.58	0.17	–0.21
T Certainty in correct answers						
<i>M (SD)</i>	68.68 (25.17)	77.46 (18.13)	81.27 (17.38)	62.66 (21.02)	63.53 (26.60)	65.71 (20.24)
Min–Max	20–100	37.50– 100	0–100	4–100	0–100	11–100
Skewness	–0.77	–0.84	–3.02	–0.60	–0.76	–0.64
Kurtosis	–0.79	–0.20	12.94	–0.02	–0.18	–0.04
T Certainty in incorrect answers						
<i>M (SD)</i>	59.01 (22.66)	71.46 (19.38)	72.80 (17.48)	57.89 (20.75)	57.47 (24.35)	61.54 (20.25)
Min–Max	23–100	37.50– 100	25–100	2.29–100	0–100	17.29– 100
Skewness	–0.12	–0.23	–0.90	–0.56	–0.54	–0.26
Kurtosis	–0.98	–1.05	1.31	0.38	–0.04	–0.65
Self-evaluated learning						

<i>M (SD)</i>	4.27 (1.33)	4.56 (1.18)	4.51 (1.28)	3.99 (1.32)	4.47 (1.05)	4.19 (0.94)
Min–Max	1–7	1–7	1–7	1–7	1–7	2–6
Skewness	–0.20	–0.36	–0.24	–0.25	–0.13	0.09
Kurtosis	0.18	1.90	0.61	0.37	1.52	0.03
Self–evaluated test performance						
<i>M (SD)</i>	3.97 (1.24)	4.22 (1.55)	4.51 (0.99)	3.43 (1.10)	3.67 (1.20)	3.52 (1.16)
Min–Max	1–7	1–7	3–7	1–6	1–7	1–7
Skewness	–0.36	–0.39	0.60	–0.30	0.15	–0.01
Kurtosis	1.31	0.01	–0.18	–0.05	0.33	0.75
Delayed testing						
Knowledge						
<i>M (SD)</i>	17 (3.74)	19.22 (3.38)	18.92 (4.62)	15 (4.60)	15.19 (5.37)	14.94 (4.21)
Min–Max	11–22	15–24	13–25	6–26	6–24	6–23
Skewness	–0.19	0.35	0.06	0.21	–0.11	0.08
Kurtosis	–1.34	–1.70	–1.62	–0.14	–0.88	–0.63
Retention						
<i>M (SD)</i>	11.30 (2.71)	12.11 (2.42)	12.17 (3.46)	9.62 (3.12)	9.81 (4.01)	9.65 (3.09)
Min–Max	5–15	9–15	7–17	4–16	3–17	4–15
Skewness	–1.23	0.24	0.13	–0.05	0.04	0.20
Kurtosis	3.05	–1.87	–1.27	–0.39	–0.68	–0.98
Transfer						
<i>M (SD)</i>	5.70 (1.95)	7.11 (1.17)	6.75 (1.60)	5.38 (1.70)	5.38 (1.98)	5.29 (1.57)
Min–Max	2–8	6–9	5–9	2–10	1–9	2–8
Skewness	–0.60	0.34	0.49	0.53	0.05	–0.02
Kurtosis	–0.29	–1.58	–1.44	0.84	–0.27	0.05
Certainty						
<i>M (SD)</i>	64.77 (19.21)	79.82 (9.77)	74.55 (10.66)	55.54 (18.35)	59.57 (21.87)	61.76 (18.57)
Min–Max	30.21– 85.69	69.31– 99.14	55.66– 91.38	5.86–83.28	25.62– 96.55	22.76– 92.10
Skewness	–1.10	0.91	–0.11	–0.93	–0.24	–0.42

Kurtosis	0.17	0.29	−0.08	0.95	−1.04	−0.44
Certainty in correct answers						
<i>M (SD)</i>	67.21 (21.08)	82.34 (10.27)	76.81 (10.58)	58.17 (20.36)	61.58 (23.28)	66.98 (19.22)
Min–Max	29.55– 87.50	68.67– 100	61.46–96	3.89–89.17	27.50– 93.33	24.55– 91.96
Skewness	−0.97	0.40	0.31	−0.81	−0.39	−0.73
Kurtosis	−0.41	−0.78	−0.34	0.46	−1.40	−0.16
Certainty in incorrect answers						
<i>M (SD)</i>	59.73 (17.68)	74.06 (10.77)	68.93 (10.23)	51.35 (16.19)	55.03 (21.53)	55.62 (17.27)
Min–Max	25.07– 81.67	64.17– 97.92	50.94– 88.33	6.75–74.17	24.91– 100	20.31– 92.57
Skewness	−1.07	1.46	0.08	−1.09	0.30	−0.01
Kurtosis	0.41	2.36	0.04	1.10	−0.67	−0.22
R Certainty						
<i>M (SD)</i>	62.85 (18.07)	78.58 (9.77)	72.79 (11.86)	53.47 (18.44)	58.49 (21.39)	59.43 (19.42)
Min–Max	30.26– 80.26	67.74– 98.68	49.74– 92.11	7.89–85.53	25–94.74	20–93.21
Skewness	−0.98	1.03	−0.23	−0.67	−0.18	−0.32
Kurtosis	−0.44	1.07	0.33	0.27	−1.06	−0.50
R Certainty in correct answers						
<i>M (SD)</i>	66.70 (19.35)	83.14 (9.05)	74.04 (14)	56.81 (20.27)	59.90 (23.15)	65.20 (21.78)
Min–Max	34.09– 83.21	72.78– 100	41.43– 96.88	8.75–89.06	25–90	20–97.73
Skewness	−1.01	0.79	−0.84	−0.58	−0.33	−0.49
Kurtosis	−0.54	−0.25	2.05	−0.26	−1.45	−0.59
R Certainty in incorrect answers						
<i>M (SD)</i>	56.34 (17.57)	70.27 (11.87)	67.76 (11.95)	48.40 (16.22)	53.30 (21.25)	51.83 (16.85)
Min–Max	25–75	56.89– 96.88	54.58–95	7.67–71.25	25–100	18.75–87
Skewness	−0.66	1.58	1.01	−0.76	0.51	−0.05
Kurtosis	−0.84	2.83	1.09	0.22	−0.48	−0.50
T Certainty						

<i>M (SD)</i>	68.41 (24.30)	82.17 (11.71)	77.89 (9.65)	59.47 (19.66)	61.62 (23.90)	66.20 (18.47)
Min–Max	27.50–96	68–100	64–93	2–92	25–100	27.50– 94.90
Skewness	–0.83	0.27	0.24	–1.13	–0.26	–0.55
Kurtosis	–0.56	–1.46	–1.01	1.65	–1.19	–0.37
T Certainty in correct answers						
<i>M (SD)</i>	69.86 (26.26)	80.90 (13.05)	81.06 (11.05)	61.13 (22.13)	63.72 (25.48)	70.49 (19.02)
Min–Max	25–95.71	62.50– 100	65.56– 98.33	0–100	25–100	29.17– 100
Skewness	–0.68	0.11	0.26	–0.71	–0.38	–0.76
Kurtosis	–1.25	–1.14	–0.89	0.92	–1.36	–0.24
T Certainty in incorrect answers						
<i>M (SD)</i>	64.04 (24.27)	85.04 (12.68)	73.01 (15.13)	56 (18.34)	58.13 (23.80)	62.41 (20.42)
Min–Max	25.17– 96.67	65–100	40–90	4–82.50	24.67– 100	25–100
Skewness	–0.51	–0.30	–1.25	–1.16	0.02	–0.03
Kurtosis	–0.99	–1.23	0.82	1.27	–1.14	–0.69
Self–evaluated test performance						
<i>M (SD)</i>	3.80 (1.03)	4.78 (1.09)	3.92 (1.16)	3.17 (1.21)	3.69 (1.35)	3.71 (1.10)
Min–Max	2–5	4–7	2–7	1–6	1–6	1–6
Skewness	–1.03	1.29	1.43	0.04	–0.23	–0.01
Kurtosis	0.49	0.77	4.99	–0.47	0	0.54

6.25 Appendix 25: ANCOVA comparisons for wood science group – Study 2

Table 139: ANCOVA comparisons on all main dependable variables in participants from educational fields related to wood science – Study 2

	ANCOVA*			Homogeneity test**		Normality test	
	<i>F</i>	<i>p</i>	η^2p	<i>F</i>	<i>p</i>	<i>W</i>	<i>p</i>
Video perception							
Pleasantness	0.02	0.977	0.00	1.96	0.146	0.97	0.022
Activation level	1.01	0.370	0.02	2.00	0.140	0.98	0.151
Differences in affective state							
Positive activation	0.10	0.907	0.00	0.09	0.918	0.97	0.030
Negative activation	1.00	0.371	0.02	0.07	0.934	0.99	0.472
Valence	0.06	0.940	0.00	0.17	0.844	0.96	0.006
Activation level ^Q	1.27	0.285		4.18	0.018	0.99	0.412
Valence	1.47	0.235	0.03	1.13	0.326	0.97	0.013
Interest and motivation							
Situational interest	0.33	0.717	0.01	0.01	0.989	0.98	0.061
Interest (delayed)	0.95	0.407	0.10	3.20	0.056	0.96	0.244
Intrinsic motivation	0.24	0.783	0.01	0.71	0.492	0.96	0.002
Learners' experience							
Paying attention	0.82	0.442	0.02	0.84	0.433	0.97	0.021
Difficulty	0.46	0.635	0.01	1.62	0.203	0.97	0.009
Exerting more effort	0.35	0.709	0.01	0.99	0.333	1.35	0.265
Enjoyment	0.79	0.459	0.02	0.46	0.630	0.96	0.002
More lessons like this	0.29	0.747	0.01	1.38	0.257	0.97	0.035
Cognitive outcomes							
Intrinsic cognitive load	0.71	0.493	0.02	0.48	0.622	0.96	0.004
Extraneous cognitive load	0.72	0.490	0.02	0.46	0.635	0.94	< .001
Germane cognitive load	0.19	0.830	0.00	0.09	0.914	0.94	< .001

Mental effort (average)	0.36	0.700	0.01	0.87	0.422	0.97	0.023
Immediate testing							
Knowledge	1.59	0.210	0.03	1.97	0.145	0.99	0.492
Retention	1.44	0.243	0.03	1.44	0.241	0.99	0.314
Transfer	1.78	0.174	0.04	0.39	0.681	0.99	0.702
Certainty	4.82	0.010	0.09	0.66	0.519	0.97	0.009
Certainty in correct answers	4.30	0.016	0.09	0.52	0.598	0.96	0.002
Certainty in incorrect answers	5.66	0.005	0.11	1.51	0.226	0.97	0.035
R Certainty	4.04	0.021	0.08	0.70	0.500	0.97	0.010
R Certainty in correct answers	3.54	0.033	0.07	0.87	0.423	0.96	0.004
R Certainty in incorrect answers	4.58	0.013	0.09	0.58	0.563	0.97	0.010
T Certainty ^Q	3.41	0.037		3.30	0.041	0.97	0.014
T Certainty in correct answers	3.24	0.044	0.07	2.84	0.063	0.97	0.028
T Certainty in incorrect answers	4.77	0.011	0.10	1.51	0.227	0.99	0.364
Self-evaluated learning	0.21	0.808	0.00	0.75	0.474	0.99	0.460
Self-evaluated test performance	1.07	0.346	0.02	1.62	0.204	0.98	0.228
Delayed testing [†]							
Knowledge	1.60	0.230	0.16	1.81	0.182	0.97	0.521
Retention	2.10	0.153	0.20	1.16	0.330	0.97	0.425
Transfer	0.30	0.744	0.03	0.47	0.630	0.97	0.534
Certainty	1.17	0.334	0.12	1.77	0.189	0.99	0.993
Certainty in correct answers	0.95	0.408	0.10	1.47	0.247	0.98	0.857
Certainty in incorrect answers	1.48	0.256	0.15	0.42	0.664	0.98	0.712
R Certainty ^Q	0.214	0.808		4.44	0.021	0.98	0.862
R Certainty in correct answers ^Q	0.17	0.844		3.33	0.050	0.98	0.767

R Certainty in incorrect answers	2.01	0.164	0.19	2.50	0.100	0.98	0.843
T Certainty	0.83	0.453	0.09	0.25	0.780	0.98	0.843
T Certainty in correct answers	0.64	0.538	0.07	0.49	0.617	0.97	0.528
T Certainty in incorrect answers	0.99	0.392	0.10	0.98	0.739	2.02	0.151
Self-evaluated test performance	0.81	0.460	0.09	0.95	0.632	0.95	0.136

Note. * $df_1 = 2$, $df_2 = 92$; ** $df_1 = 2$, $df_2 = 103$; † $df_1 = 2$, $df_2 = 17$; ** $df_1 = 2$, $df_2 = 28$; R – retention, T – transfer, ^Q – Quade test results reported instead of ANCOVA due to the homogeneity of variances assumption not being met ($df_2 = 103/28$); covariates included are prior tested knowledge, prior interest, LexTALE score, the five personality characteristics, and PANAVA–KS baseline measures, except for the Activation level and Valence variables, which had the activation level and valence baseline measures

6.26 Appendix 26: ANCOVA comparisons for non-wood science group – Study 2

Table 140: ANCOVA comparisons on all main dependable variables in participants from educational fields not related to wood science – Study 2

	ANCOVA*			Homogeneity test**		Normality test	
	<i>F</i>	<i>p</i>	η^2p	<i>F</i>	<i>p</i>	<i>W</i>	<i>p</i>
Video perception							
Pleasantness	6.55	0.002	0.07	0.95	0.387	0.99	0.151
Activation level	1.86	0.159	0.02	0.14	0.870	0.99	0.046
Differences in affective state							
Positive activation ^Q	1.01	0.368	0.01	5.01	0.008	0.99	0.705
Negative activation	2.18	0.116	0.02	0.99	0.374	0.99	0.054
Valence	2.64	0.074	0.03	0.83	0.436	0.98	0.002
Activation level	2.09	0.127	0.02	2.16	0.118	0.97	0.002
Valence	4.75	0.010	0.05	2.80	0.063	0.98	0.017
Interest and motivation							
Situational interest	4.74	0.010	0.05	0.98	0.378	0.99	0.404
Interest (delayed)	1.05	0.356	0.03	1.57	0.214	0.99	0.761
Intrinsic motivation	3.54	0.031	0.04	0.18	0.833	1.00	0.883
Learners' experience							
Paying attention	5.30	0.006	0.06	0.23	0.796	0.99	0.143
Difficulty	0.09	0.911	0.00	0.59	0.554	0.98	0.006
Exerting more effort	0.21	0.811	0.00	0.13	0.232	0.99	0.232
Enjoyment	2.18	0.116	0.02	1.42	0.244	0.99	0.398
More lessons like this	4.70	0.010	0.05	1.25	0.288	0.99	0.645
Cognitive outcomes							
Intrinsic cognitive load	0.04	0.957	0.00	0.73	0.482	0.98	0.021
Extraneous cognitive load	1.40	0.250	0.02	1.45	0.238	0.97	< .001
Germane cognitive load	1.48	0.230	0.02	0.62	0.538	0.98	0.027

Mental effort (average)	1.47	0.234	0.02	1.80	0.169	0.99	0.676
Immediate testing							
Knowledge	0.32	0.726	0.00	0.36	0.701	0.99	0.126
Retention	0.50	0.610	0.01	1.11	0.332	0.99	0.513
Transfer	0.06	0.942	0.00	1.91	0.152	0.99	0.045
Certainty ^Q	0.03	0.873		4.98	0.008	0.98	0.011
Certainty in correct answers ^Q	0.42	0.518		5.18	0.006	0.98	0.006
Certainty in incorrect answers ^Q	0.17	0.678		3.90	0.022	0.99	0.087
R Certainty ^Q	0.00	0.954		4.03	0.019	0.99	0.239
R Certainty in correct answers ^Q	0.59	0.443		4.49	0.012	0.99	0.043
R Certainty in incorrect answers	0.06	0.943	0.00	2.41	0.093	0.99	0.543
T Certainty ^Q	0.23	0.798	0.00	2.87	0.059	0.97	0.001
T Certainty in correct answers	0.24	0.784	0.00	2.77	0.065	0.98	0.009
T Certainty in incorrect answers ^Q	0.75	0.389		3.21	0.043	0.99	0.358
Self-evaluated learning	3.84	0.023	0.04	2.76	0.066	0.98	0.002
Self-evaluated test performance	1.11	0.332	0.01	0.09	0.916	0.99	0.455
Delayed testing [†]							
Knowledge	0.94	0.394	0.03	2.21	0.117	0.98	0.238
Retention	1.02	0.367	0.03	0.84	0.437	0.97	0.084
Transfer	0.46	0.631	0.01	1.00	0.374	0.99	0.809
Certainty	1.75	0.181	0.05	0.50	0.611	0.96	0.018
Certainty in correct answers	1.97	0.147	0.05	0.58	0.562	0.96	0.007
Certainty in incorrect answers	1.36	0.264	0.04	1.00	0.372	0.99	0.622
R Certainty	2.03	0.139	0.05	0.20	0.816	0.97	0.038
R Certainty in correct answers	1.87	0.161	0.05	0.28	0.754	0.96	0.008

R Certainty in incorrect answers	1.50	0.230	0.04	0.76	0.471	0.99	0.946
T Certainty	1.22	0.301	0.03	1.07	0.346	0.98	0.179
T Certainty in correct answers	1.54	0.220	0.04	1.02	0.365	0.98	0.307
T Certainty in incorrect answers	1.10	0.340	0.03	1.08	0.345	0.99	0.499
Self-evaluated test performance	3.35	0.040	0.08	0.02	0.980	0.99	0.978

Note. * $df_1 = 2$, $df_2 = 179$; ** $df_1 = 2$, $df_2 = 190$; † * $df_1 = 72$, $df_2 = 83$; ** $df_1 = 2$, $df_2 = 28$; R – retention, T – transfer, ^Q – Quade test results reported instead of ANCOVA due to the homogeneity of variances assumption not being met ($df_2 = 191$); covariates included are prior tested knowledge, prior interest, LexTALE score, the five personality characteristics, and PANAVA–KS baseline measures, except for the Activation level and Valence variables, which had the activation level and valence baseline measures