Lídia Vinczéné Fekete

Student Engagement and Emotions in Higher Education

Affective Patterns and Boredom

During Instructional Video Viewing

Corvinus University of Budapest

Doctoral School of Sociology and Communication Science

Communication Science Doctoral Program

Supervisor: Dr. Judit Gabriella Sass

Copyright © Lídia Vinczéné Fekete

Corvinus University of Budapest

Doctoral School of Sociology and Communication Science

Communication Science Doctoral Program

Student Engagement and Emotions in Higher Education

Affective Patterns and Boredom

During Instructional Video Viewing

Doctoral dissertation

Lídia Vinczéné Fekete

Budapest, 2026

Table of contents

1 INTRODUCTION 15

1.1 Context and relevance 15

1.2 Research gaps 16

1.3 Research questions and analytic expectations 19

1.4 Scope 20

1.5 Theoretical foundations: overview 21

1.5.1 Epistemic and ontic orientation of the study 21

1.5.2 Communication 24

1.5.3 Education Science 25

1.5.4 Psychology 26

1.5.5 Affective Computing 27

1.6 The structure of the thesis 27

2 STUDENT ENGAGEMENT 29

2.1 Definitions 29

2.2 The evolution of the concept 30

2.3 Conceptualisation 32

2.3.1 Distinguishing student engagement from motivation 33

2.3.2 Is disengagement the opposite of engagement? 33

2.4 Focus of the thesis related to student engagement 33

3 ACADEMIC EMOTIONS 35

3.1 Emotions 36

3.2 Emotional taxonomies 37

3.2.1 Categorical taxonomies 37

3.2.2 Dimensional taxonomies 39

3.3 The role of academic emotions 40

3.3.1 The Control-Value Theory 40

3.3.2 Emotions predominantly positively associated with engagement and learning 44

3.3.3 Emotions with mixed or context-dependent associations with engagement and learning 44

4 BOREDOM 47

4.1 A growing field of research 47
4.2 Conceptualisation and theoretical accounts of boredom 48
4.2.1 Definitions 48
4.2.2 State and trait boredom 49
4.2.3 Academic boredom 50
4.2.4 Conceptual boundaries 51
4.2.5 Theoretical accounts of boredom 53
4.2.6 Chosen theoretical framework and working definition of boredom 55
4.3 Unresolved theoretical questions 55
4.3.1 Is boredom an emotion? 55
4.3.2 Does boredom always involve low arousal? 56
4.3.3 What makes an activity boring: difficulty, lack of meaning, or both? 58
4.3.4 Two pathways for alleviating boredom: interest and enjoyment 60
4.4 Strategies to reduce boredom 61
4.4.1 Coping with boredom as a student 61
4.4.2 Instructional interventions to prevent and reduce boredom 63
5 THE POTENTIAL OF INSTRUCTIONAL COMMUNICATION 66
5.1 Rhetorical and relational perspectives in instructional communication 66
5.2 Teacher immediacy and clarity 71
5.2.1 Immediacy 71
5.2.2 Clarity 72
5.2.3 Control and value as learner-side appraisals 73
5.3 Instructional communication fostering cognitive and affective engagement in online learning 73
6 INSTRUCTIONAL VIDEOS 76
6.1 Factors influencing the effectiveness of instructional videos 76
6.2 The Cognitive Load Theory and the Cognitive Theory of Multimedia Learning 77
6.3 Connecting multimedia learning and instructional communication 79
6.4 A conceptual mapping tool linking multimedia learning and instructional communication 80
7 MEASUREMENT OF AFFECTIVE RESPONSES 84
7.1 Detecting and measuring emotions and engagement 84
7.2 The relationship between achievement emotions and student engagement 85

7.3 Signs of engagement and disengagement 85
7.4 Methods applied to measure engagement 88
7.4.1 Subjective methods 91
7.4.2 Objective methods 93
7.5 Automated facial expression analysis 94
7.6 Methodological implications of multimodal measurements 96
8 METHODOLOGY 98
8.1 Framework of research design 98
8.2 Sample 100
8.3 Stimulus 101
8.4 Measurement instruments 103
8.4.1 Questionnaire 103
8.4.2 Interview 103
8.4.3 Facial analysis 104
8.5 Procedure 104
8.6 Ethical considerations 105
8.7 Data analysis 106
8.7.1 Coding-based qualitative content analysis of instructional video 106
8.7.2 Self-reported experiences 109
8.7.3 Facial expression data 110
9 RESULTS 116
9.1 Application of the coding framework to the instructional video 116
9.2 RQ1: Design features most strongly associated with boredom 118
9.2.1 Raw mean values of boredom and interest 118
9.2.2 Relative frequency analysis of boredom peak moments 121
9.2.3 Associations of multimedia learning features with boredom and neutral indicators 124
9.2.4 Students' self-perceived boredom - Interviews 128
9.3 RQ2: Co-occurrence between design features and affective response patterns 133
9.3.1 Insight into some selected relationships between design features and basic emotions 134
9.3.2 Overview of main patterns 135
9.4 Alignment across data sources 137

9.4.1 Congruent results 137
9.4.2 Incongruent results 138
10 DISCUSSION 144
10.1 Boredom in the context of instructional video viewing 144
10.2 Instructional communication, multimedia features and affective responses 145
10.3 Interpreting incongruent findings 146
10.3.1 Relationship between mental states and bodily signals 147
10.3.2 Reliability of identifying affective-cognitive changes from facial data 147
10.3.3 Co-occurrence of multiple affective states and overlaps between their expression 147
10.3.4 Negative affectivity during an activity is not equal to disengagement . 149
10.4 RQ3: Alignment across data sources and methods 150
10.5 Implications for practice 150
10.5.1 Teaching practice 150
10.5.2 Instructional videos 152
10.5.3 Implications for other fields 158
10.6 Limitations and directions for future research 159
11 CONCLUSION 162
11.1 Summary of aims and key findings 162
11.2 Main contributions and novel insights of the thesis 164
11.2.1 Theoretical contribution to understanding boredom 164
11.2.2 Conceptual contribution: connecting multimedia learning and instructional communication 165
11.2.3 Methodological contribution 167
11.3 Concluding remarks 169
12 REFERENCES 172
13 AI DECLARATION 188
14 APPENDICES 189
Appendix A: Differences between student motivation and engagement 189
Appendix B: Overview of the context, methods and findings of studies applying FaceReader (FR) 190
Appendix C: Analysis of ethical risks and benefits 196
Appendix D: Student questionnaires 198

Appendix E: Screenshot of the FaceReader 9.1 interface (Noldus Information Technology) 206
Appendix F: Effects of multimedia learning features on facial expressions related to boredom 207
Appendix G: Effects of multimedia learning features on facial expressions related to basic emotions 209
Appendix H: Interview questions..... 212

List of Figures

Figure 1: Disciplinary and conceptual positioning of the thesis 23
Figure 2: Delimitation of dissertation topics based on relevance and research gaps. 28
Figure 3: Reciprocal causation of academic emotions, engagement, and their antecedents and outcomes. 35
Figure 4: The circumplex model of affect developed by Russell. 39
Figure 5: Annual number of publications on academic boredom indexed in the Scopus database 47
Figure 6: Trait boredom and the meaning and attention model. 60
Figure 7: Antecedents and consequences of boredom. 65
Figure 8: A four-dimensional model of rhetorical and relational traditions in instructional communication 69
Figure 9: Distribution of emotional states by affective measurement channels. 90
Figure 10: Conceptual framework and analytical steps for detecting emotional responses and modelling re-engagement strategies. 99
Figure 11: Procedure of the observation 105
Figure 12: Duration of certain multimedia principles applied in the instructional video. 117
Figure 13: Changes in the mean raw values of boredom and interest while viewing a 22.6-minute video presentation 119
Figure 14: Boredom peak phases per 10-second time windows (N=49) 123
Figure 15: Associations between selected multimedia-learning features and FaceReader-labelled boredom..... 125

Figure 16: Attention-supporting communication features mentioned by participants 128
Figure 17: Attention-impairing communication features mentioned by participants 129
Figure 18: Segmentation of the video according to boredom phases 131
Figure 19: Changes in the mean level of arousal and valence before and after the video segment deemed most boring by students 132
Figure 20: Individual trajectories of FaceReader-derived boredom before, during, and after the video segment most frequently identified as boring 133
Figure 21: Aggregated mean values of students' emotions on facial expressions (N=50). Noldus FaceReader 9.1 data. 134
Figure 22: Distribution of dominant expressions, excluding "neutral" (N=50). Noldus FaceReader 9.1 data. 134
Figure 23: Summary heatmap of the relationship between affective states and the application of certain multimedia principles 136
Figure 24: Mean level of sadness, during periods of self-reported interest and boredom 139

List of Tables

Table 1: The three-dimensional taxonomy of achievement emotions 42
Table 2: The effect of emotions on academic engagement, motivation and performance (see the sources below the table)..... 43
Table 3 - Profiles of Boredom and Their Predicted Outcome as a Function of Meaning and Attention..... 58
Table 4: Selected concepts reflecting rhetorical and relational emphases in instructional communication..... 68
Table 5: Empirical support for instructional-video design principles identified by Fyfield et al. (2022)..... 78
Table 6: Illustrative areas of overlap and divergence between instructional communication variables and multimedia-learning principles 79
Table 7: Conceptual mapping of multimedia-learning principles and coded instructional-video features onto rhetorical and relational communication functions81
Table 8: Nonverbal signs of engagement and disengagement..... 87
Table 9: Categorisation of measurement methods for learning engagement..... 91
Table 10: Segment-level overview of the instructional video stimulus and its instructional strategies and multimedia features..... 101
Table 11: Data collection and data analysis methods 106
Table 12: Coding framework for the qualitative analysis of instructional video (1-second intervals)..... 107
Table 13: Video features associated with congruent patterns of viewer boredom and interest in the 30-second time-window analysis (N = 25) 120
Table 14: Features of phases without boredom peak moments identified through relative frequency analysis..... 122
Table 15: Assumption test for normality for the calculation of the T-test..... 124
Table 16: Associations between selected multimedia-learning features and FaceReader-labelled boredom (paired-samples t-tests and Wilcoxon signed-rank tests) 126
Table 17: Number and proportion of participants mentioning immediacy and clarity-related features perceived as supporting their attention. 129
Table 18: Facial expression identified as sadness, during periods of self-reported interest and boredom..... 138

Table 19: Effect of certain multimedia principles on “disgusted” as measured by FaceReader, using the Wilcoxon signed rank test (JASP (Version 0.19.2)). 140
Table 20: Assumption test for normality applied for the calculation of the t-test comparing emotions measured during self-reported boredom and interest. 141
Table 21: Intensity and level of affective states during self-reported boredom and interest periods 141
Table 22: Correlation between the facial expression of 'boredom', emotional and cognitive interest, and changes in positive and negative affectivity after video watching. 142

Acknowledgements

My deepest thanks go to my supervisor, Judit Sass, whose expertise, careful judgement, and trust have supported this research throughout the years. She is also an important professional role model for me as both a teacher and a researcher, embodying the balance between academic and relational support that has inspired me to engage more deeply with the study of teaching and learning.

Special thanks to Tamás Bokor, whose instructional video and students gave me the privilege of gaining insight into his teaching practice and examining one small aspect of it in considerable depth. I am so grateful to Alexandra Béni, Judit Gáspár, Rita Polt, and Richárd Szántó for giving me a glimpse into their inspiring teaching practices. My thanks extend to all students who participated in the study and generously contributed their time and perspectives. I am especially grateful to Péter Soltész for his methodological guidance and careful review of the statistical analyses, which gave me confidence in the rigour and reliability of my analytical work.

I also wish to thank the Doctoral School of Sociology and Communication Science and the colleagues I met there, for the intellectual community that became an important part of these years. Jessie Labov and Réka Benczes gave me something particularly valuable: high expectations accompanied by equally high levels of support, care, and confidence in me and my research.

I owe a great deal to my colleagues, who stood beside me throughout this long process. I particularly want to thank Éva Bodnár, Olga Csillik and Katalin Dósa, whose encouragement, flexibility and support as my workplace leaders made it possible for me to pursue this research alongside my professional responsibilities.

Finally, my most personal thanks go to those closest to me, who have lived through this journey with me: my friends, Eszter and Edit; my sister, Judit; my parents; my husband, Gábor; and my children. They shared both the difficult periods and the achievements, stood by me throughout, and made me feel that I could always count on them – sometimes through encouragement, sometimes through a timely word of caution. Most importantly, they made me feel that I mattered to them, and that what mattered to me mattered to them, too. This journey has asked something of all of us, and in that sense, this degree will be a shared achievement.

1 INTRODUCTION

"…*achievement emotions relate to performance, over and above the influence gender, age, academic ability, and personality traits*."
(Pekrun et al. 2023, p. 172)

"*the difference between knowing and teaching is communication*"
(Hurt, Scott and McCroskey 1978, p. 3)

Student engagement is the concept that guided the research journey leading to the present thesis. Questions concerning this state of goal-directed, active and focused involvement in learning gradually gave rise to further questions, directing attention towards related fields of study and concepts. This chapter outlines the progression of the topics addressed in the thesis and identifies the research gaps. It also maps the path leading to the research questions, showing how these gaps informed the methodological design, the selection and collection of data, and, subsequently, the analysis of the data.

1.1 Context and relevance

Higher education is being reshaped by technological change, financial pressures, evolving student needs, and the ongoing expansion of flexible and blended learning. The rapid shift to online education during the COVID-19 pandemic accelerated developments that were already underway and made both the possibilities and the limitations of mediated teaching more visible. At the same time, higher education faces growing financial constraints, intensifying competition from AI-driven and market-led training programmes, and declining public perception of its value. Universities are therefore increasingly challenged to create learning experiences that students perceive as meaningful, engaging, and relevant to their needs. These expectations are consistent with the broader educational direction represented by the OECD Learning Compass 2030, which advocates a holistic and student-centred approach extending beyond knowledge acquisition and academic performance to individual and collective wellbeing (OECD 2019). From this perspective, the quality of the learning experience matters not only because it supports effective learning, but also because fostering individual and collective wellbeing is itself a fundamental, long-term purpose of education.

Research on student engagement has developed rapidly over the past two decades, accompanied by growing interest in academic emotions. The pandemic intensified attention to an already developing 'relational turn' in education (Matthews et al. 2024), while research on belonging, relationship-rich education and need-supportive instructional

communication has shown that these values are associated with students' motivation, persistence, achievement, retention, and wellbeing (Blake, Capper and Jackson 2022; Kift 2024; Mercer-Mapstone et al. 2024). The growing research body on academic emotions exposed that affective factors are more influential than previously thought (Matthews, Lodge and Johnstone 2024; Vaughan et al. 2023): they influence attention, motivation, strategy use, persistence, and performance, while learning experiences and outcomes, in turn, shape students' emotions (Pekrun 2006; Pekrun and Linnenbrink-Garcia 2022; Pekrun, Marsh, Elliot, et al. 2023). Negative deactivating emotions, particularly *boredom* and hopelessness, show especially strong adverse associations with learning strategies, performance, and wellbeing (Pekrun, Marsh, Elliot, et al. 2023).

The problem becomes especially salient in *asynchronous video*-based learning. The shift to online learning during the pandemic revealed that the strategies, such as eye contact, nonverbal behaviours that communicate teacher immediacy and students' implicit messages about their needs (Richmond, Gorham and McCroskey 1987; Mazer 2022) were *unavailable* in the online space (Dixson et al. 2017), especially in asynchronous settings. Following the pandemic, face-to-face learning returned to normal. Still, the opportunities and challenges presented by instructional videos have remained, raising questions and prompting instructors and curriculum developers to develop innovations, theoretical ideas, and recommendations (see Chapter 6). Meanwhile, maintaining students' attention and engagement during instructional videos remains challenging, with boredom and isolation representing potential difficulties in asynchronous learning environments.

1.2 Research gaps

The central problem addressed by this thesis concerns how verbal and nonverbal teacher behaviours and instructional design features can shape learners' emotional responses in ways that support learning, particularly in the context of instructional videos.

The research gaps addressed in this thesis emerged progressively through the synthesis of several partly distinct bodies of literature, each illuminating a different aspect of the central problem. Together, these gaps approach the central problem from complementary perspectives. The following chapters trace the theoretical path through which these interconnected gaps were identified, while the present section provides a brief overview of the thesis's rationale and direction.

The first gap concerns the understanding of *boredom* during learning. Although academic boredom is recognised as an important state closely related to learning and engagement, fundamental questions remain about what gives rise to boredom and how learners might overcome or alleviate it. Competing and complementary accounts emphasise attentional mismatch, insufficient meaning, control and value, while boredom may also vary substantially in its arousal characteristics and temporal development (Goetz et al. 2024; Tam et al. 2021; Westgate and Wilson 2018). Further questions concern the function of boredom in learning, the extent to which it may be adaptive rather than merely detrimental, and how responsibility for its regulation is distributed between learners and instructors. This also raises the practical question of which strategies may be most relevant to preventing boredom or supporting re-engagement once it occurs. These questions, examined in detail in Chapter 4, create a need for further empirical evidence on the situational conditions associated with boredom and on how boredom develops and changes over the course of an ongoing learning activity. A broader question to which the thesis seeks to contribute concerns the *role of unpleasant affective states* in learning, including not only boredom but also the possible functions of other negatively valenced experiences.

A second unresolved issue concerns the *measurement* of learners' affective states. Research on academic emotions still relies heavily on self-report measures, which provide essential access to learners' subjective experience but offer limited temporal resolution and cannot capture several aspects of affective responding. Automated measures, including facial expression analysis, offer the possibility of fine-grained and continuous observation (D'Mello, Dieterle and Duckworth 2017; Yadegaridehkordi *et al.* 2019; Pekrun, Marsh, Elliot, *et al.* 2023), but the psychological interpretation of their outputs remains uncertain. There is therefore a need not only to apply such methods in authentic learning contexts, but also to systematically compare automated indicators with learners' own accounts and examine where different measurement channels converge and diverge (Landmann 2023).

A third underexplored area concerns *multimedia learning research* and, more specifically, the range of instructional design features represented in its established design principles. Multimedia learning research has established a substantial evidence base on design principles that have also been adapted to instructional video studies (Fyfield, Henderson and Phillips 2022; Mayer 2021). However, these principles have been developed primarily

around cognitive processing, while comparatively few explicitly address instructional design elements linked to immediacy and cues that foster a sense of personal connection with learners. Although multimedia learning research has increasingly recognised the relevance of affective and emotional design (Schrader, Kalyuga and Plass 2021), these dimensions remain underrepresented in its established design principles. This imbalance is also reflected in empirical research, where instructional video studies have predominantly examined cognitive and performance-related outcomes, leaving considerably less evidence on how specific design features are associated with learners' moment-to-moment affective responses, including boredom (Fyfield, Henderson and Phillips 2022).

A further conceptual and operational limitation concerns the operationalisation of *instructional communication* for the systematic analysis of instructional materials. Instructional communication research provides a rich body of constructs describing rhetorical and relational teacher communication, including clarity and immediacy. However, the variables and behaviours examined in this field do not constitute an integrated category system that can be readily operationalised to systematically code instructional stimuli. As mentioned above, multimedia learning principles provide more directly observable design categories, but they include relatively few clearly differentiated and sufficiently specific design features that capture the relational functions of teacher communication. As a result, it remains difficult to analyse observable instructional features in a way that also reflects their broader relational and rhetorical communicative functions.

The gaps outlined above emerge from partly distinct theoretical and methodological traditions, but they converge on the central problem introduced earlier, each from a complementary perspective.

1. From the *learner* perspective, the emergence, development, and regulation of *boredom* and other affective states;
2. from the *researcher's* perspective, their *measurement* and interpretation;
3. and from the *instructor's* and *instructional designer's* perspectives, the design and communication features that may shape learners' affective responses.

The empirical study brings these strands together by examining how selected instructional video features co-occur with learners' affective responses and by comparing patterns across automated facial analysis, self-report, and interview data.

1.3 Research questions and analytic expectations

In line with the aims discussed in Section 1.2, the thesis addresses the following research questions:

RQ1: *Which design features were associated with student boredom within the analysed segments of the observed instructional video?*

RQ2: *How do immediacy- and clarity-related design features co-occur with learners' affective response patterns during instructional video viewing, as indicated by automated facial expression analysis?*

RQ3: *To what extent do automated facial expression indicators converge with or diverge from questionnaire- and interview-based accounts of boredom and interest during instructional video viewing?*

Based on the literature review and the preliminary observations conducted prior to the present research, the following analytic expectations have been formulated in relation to the above questions:

AE1: *Within the analysed segments of the observed instructional video, design features identified in the literature as attention-supporting are expected to be associated with lower levels of student boredom.*

AE2: *Immediacy- and clarity-related features are expected to be associated with more positive valence, higher arousal, higher levels of FaceReader-labelled happy expression, and lower levels of FaceReader-labelled negative expressions than segments in which these features are absent.*

AE3: *Partial convergence and theoretically informative divergence are expected because automated facial analysis, questionnaires and interviews capture different components, temporal levels and modes of expression of affective experience.*

1.4 Scope

The thesis aims to identify behavioural factors in instructional communication and further features of instructional videos that are associated with student learning. It nevertheless does not attempt to examine the engagement construct in its entirety; rather, it focuses on indicators of its affective and attentional dimensions. Within this scope, based on analyses of relevant theories, empirical studies, and empirical measurements of student emotions, it examines the relationships between instructional communication and learners' affective states, including valence, arousal, and basic emotions.

The affective state of *boredom* received a central focus, due to its crucial role in student engagement. The thesis aims to contribute to a better understanding of its possible antecedents and causes, considering the controversies and questions raised by several studies (see Section 4.3).

The empirical research discussed relies on multimodal data. A methodology review of measurement strategies was conducted to ensure that the analysis, comparison, and interpretation of these results were consistent with the practices reported and the conclusions drawn in previous empirical research. The methodological insights gained from this study's data analysis can support ongoing research into how academic emotions are identified and measured, particularly through the interpretation of nonverbal behaviours observed during instructional video viewing.

Rather than seeking to establish direct causal effects, the thesis aims to identify patterns in how certain instructional video elements are associated with students' affective responses. These insights may support the development of pedagogical strategies that strengthen the relational aspects of communication and help instructors design video content that fosters attention, engagement, and a sense of connection in online learning environments.

Thus, while the observations were conducted in the specific context of individual, asynchronous, online learning, the implications of the findings extend beyond this setting. Several identified patterns – such as the associations with visual design, personalisation, or storytelling on engagement – may also be relevant to traditional and blended learning environments and warrant examination in these contexts. As such, the thesis seeks to

provide practical insights to inform everyday teaching practices in higher education, particularly in designing learning experiences that support emotional and cognitive engagement, regardless of the delivery mode.

The current thesis draws on *educational psychology* and *behavioural research* to examine student emotions and their measurement; on *communication science*, especially interpersonal and relational communication, to analyse the effects of teacher communication; and on *instructional and media communication*, focusing on multimedia-based teaching tools, methods, and principles. Its practical implications are grounded in *educational science* and interpreted in relation to major theories of teaching and learning. The dissertation also engages with *affective computing*, linking computer science, communication science, and psychology. Automated facial expression analysis is used not only as a method, but also examined as a methodological focus and contribution, reflecting growing recommendations for fine-grained, automated measurement of affective states (Pekrun 2006; Sinatra, Heddy and Lombardi 2015; D'Mello, Dieterle and Duckworth 2017; Goetz *et al.* 2024).

1.5 Theoretical foundations: overview

This section sets out the main ontic and epistemic commitments that guide the thesis and explains how they are reflected in its disciplinary perspectives, theoretical choices, methodological design, and interpretation of empirical evidence.

1.5.1 Epistemic and ontic orientation of the study

The research discussed in the thesis is primarily situated within instructional communication, a branch of communication science that provides the main disciplinary perspective from which the research problem and research questions are formulated. The purpose of this section is to clarify this primary theoretical orientation and to identify the neighbouring disciplines that inform the research where the phenomena under investigation extend beyond the boundaries of communication science.

Alongside this primary disciplinary orientation, the work draws selectively on constructivist perspectives to articulate its assumptions about learning, interpretation, and the nature of the evidence generated by different methods. Constructivism itself comprises a family of positions rather than a single unified doctrine (Phillips 1995), and its meaning

therefore needs to be specified at the level at which it is applied. These epistemic assumptions also shape what the present thesis treats as evidence.

Learners' engagement and emotions cannot be observed completely or directly through any single measurement channel. Self-reports provide access to consciously interpreted experience; interviews provide retrospective accounts and contextual meaning; facial data provide temporally fine-grained indicators of observable facial activity; and analysis of the instructional stimulus provides information about the communicative and design conditions with which these responses co-occur. Each type of evidence, therefore, provides access to a different aspect of the phenomenon and is shaped by its own measurement system, theoretical assumptions, temporal resolution, and possible sources of error.

The multimodal design follows from this epistemic position. Agreement between measures can strengthen an interpretation, but disagreement is not automatically treated as a measurement failure. Different methods may diverge because experienced, consciously interpreted, remembered, expressed, and automatically classified responses are not equivalent. The function of multimodal evidence in this thesis, therefore, extends beyond triangulation understood as confirmation. Its function also includes complementarity and the identification of theoretically informative discrepancies. Greene, Caracelli and Graham (1989), for example, distinguish triangulation from complementarity, initiation, development, and expansion as separate purposes of integrating qualitative and quantitative evidence.

This distinction is particularly important for automated facial analysis. Facial movements constitute observable data; emotion categories produced by automated analysis constitute model-based classifications of these movements. They are not treated as direct observations of internal emotional states. Research on emotional expression shows substantial variation in the relationships between facial movements, situations, cultures, and experienced emotions and cautions against assuming one-to-one mappings between particular facial configurations and discrete internal states (Barrett et al. 2019). Accordingly, terms such as “happy”, “sad”, “angry”, and “bored” are treated as software-generated labels when they refer to FaceReader outputs. Their psychological interpretation depends on the convergence or divergence of these outputs with other indicators and with the instructional context.

Figure 1 maps the main theories and frameworks according to both their primary disciplinary grounding and their conceptual relevance to the present research.

In educational science, constructivism concerns the learner's active role in meaning-making: learning cannot be reduced to information transfer, as learners actively interpret and organise what they encounter (Biggs and Tang 2011). In psychology, the thesis does not adopt a single constructivist theory of emotion as its ontological foundation. It instead draws on theories that assign explanatory importance to appraisal, attention, meaning, goals, and context, while recognising ongoing disagreement over how emotion categories should be conceptualised (Ekman 1992; Barrett 2017). In communication science, the relational commitment concerns the constitutive and relational role of communication.

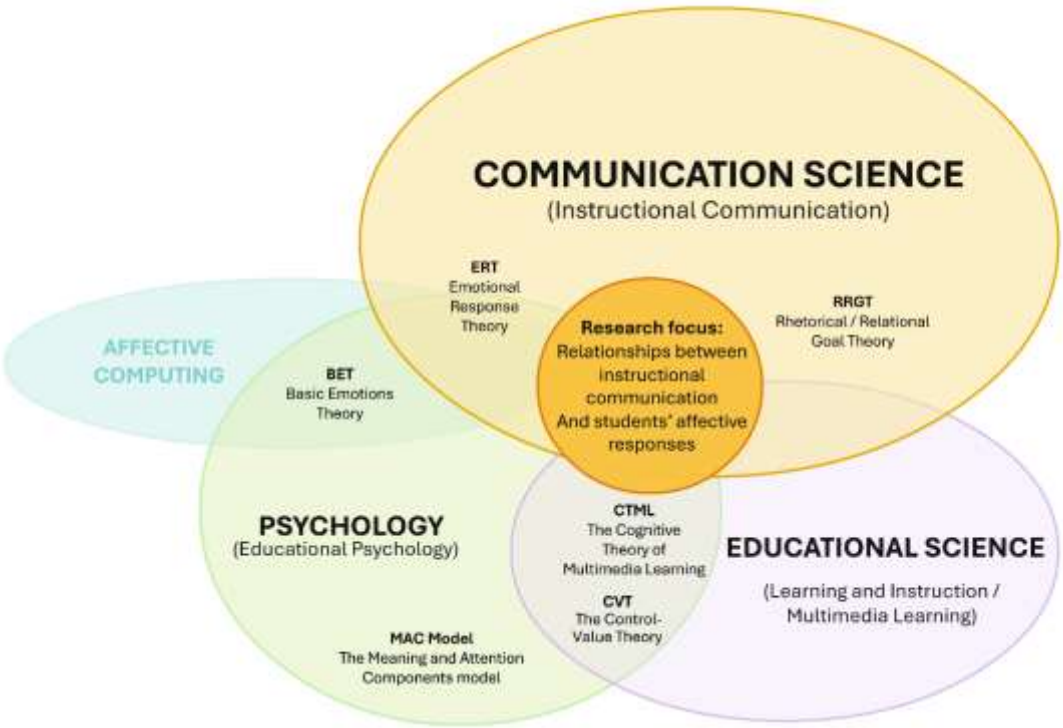


Figure 1: Disciplinary and conceptual positioning of the thesis

Theories drawn from other disciplines are used according to the explanatory problems for which they were developed. The Control-Value Theory of Achievement Emotions explains how learners' appraisals of control and value are related to achievement emotions. The Meaning and Attention Components model addresses attentional fit and perceived

meaning in boredom. The Cognitive Theory of Multimedia Learning identifies characteristics of instructional design relevant to cognitive processing. These theories operate at different conceptual and explanatory levels and are therefore treated as potentially complementary rather than necessarily competing.

These commitments also delimit the claims that can be drawn from the empirical study. The research does not seek to establish context-independent causal laws concerning individual instructional features. It examines theoretically meaningful associations, co-occurrence patterns, temporal changes, and convergence or divergence across indicators within a defined instructional context. Statistical procedures assess the strength, consistency, and uncertainty of patterns in the observed data; they do not turn observational contrasts into evidence of causation. Qualitative accounts provide information about how learners interpreted specific episodes, while cross-method comparisons examine whether different indicators yield compatible or divergent representations of the same instructional periods.

Evidence in the thesis is therefore situated but not arbitrary. Interpretations are constrained by the collected data, statistical analyses, established theoretical expectations, and known limitations of the measurement procedures, while remaining tied to the sample, stimulus, educational setting, and operational definitions from which they were generated. Recurrent patterns across participants, temporal segments, indicators, or methods can support theoretically meaningful propositions, but the level of generalisation must remain proportionate to the research design. Practical implications are consequently formulated as context-sensitive possibilities and research-informed design principles rather than universally applicable prescriptions.

1.5.2 Communication

The thesis is situated primarily within communication science and, more specifically, instructional communication. This disciplinary grounding shapes the questions asked about teaching and learning: instructional features are examined as communicative choices that shape learners' relations with the instructor and the learning process.

The broader perspective of this approach is compatible with Celeste Condit's (2006) *relational account* of communication, in which relating is treated as a fundamental function of communication and relationships, and participants are continually constituted through communicative processes. In Craig's (1999) metamodel, communication theory is organised into seven traditions that emphasise different primary functions and assumptions of communication, including rhetorical, sociocultural and other perspectives. The present approach draws particularly on the rhetorical tradition's focus on purposeful communicative action and the sociocultural tradition's view of communication as constitutive of social relationships and shared realities.

Instructional communication itself combines rhetorical and relational concerns. The Rhetorical/Relational Goal Theory proposes that instructors pursue both rhetorical and relational goals and that learners differ in the relative importance of their academic and relational needs (Mottet, Frymier and Beebe 2006). This makes the theory particularly compatible with the context-sensitive position of the present thesis: communication strategies cannot be assumed to have identical consequences for all learners. Emotional Response Theory (Mottet, Frymier, and Beebe, 2006) provides a more specific explanatory model that links instructional communication, emotional response, and approach-avoidance tendencies. It is used here as a theoretically informative model rather than as an established causal sequence, given that subsequent empirical testing found only partial support for its proposed causal structure (Horan, Martin and Weber 2012).

Within this framework, rhetorical and relational communication are treated as complementary orientations. Rhetorical goals and clarity primarily concern the intelligibility, organisation, and cognitive accessibility of instructional messages, whereas relational functions and immediacy refer to communicative behaviours that reduce perceived psychological distance and support interpersonal connections. Their empirical manifestations may overlap, and the same behaviour can serve both functions depending on context.

1.5.3 Education Science

Educational science provides the context in which the communication problem examined in the thesis acquires its purpose. The relational perspective on communication aligns closely with the *constructivist theory of learning*, which holds that knowledge cannot be

transmitted but is constructed individually by the learner. According to this approach, the teacher's main responsibility is to facilitate learning by creating an environment that supports autonomy, self-direction, and meaningful engagement principles that are especially critical in online learning contexts (Biggs and Tang 2011).

This educational perspective is also compatible with the recent models of student engagement that conceptualise engagement as dynamic and situated at the interface between students and their educational environments (Kahu and Nelson 2018). The present thesis focuses on micro-level processes at this interface, particularly the relationships among instructional communication, learners' affective responses, attention, and engagement. This focus does not imply that instructors alone determine engagement; student characteristics and broader institutional and sociocultural conditions form an important part of the context within which instructional communication operates. According to Condit's relational approach to the educational context (Condit 2006), the aim of communication is to actively constitute the relationship between the affected parties, in this case, between the teacher and the student. In this view, emotions and engagement emerge not in isolation, but as products of these relational processes – even in asynchronous, video-based instruction.

While the present research gives particular attention to the affective dimensions of learning, the Cognitive Theory of Multimedia Learning (CTML) complements this perspective by providing a more specific account of the *cognitive processes* involved in learning from instructional messages. CTML explains how learners select, organise, and integrate verbal and visual information within a limited-capacity cognitive system (Mayer 2021), but its particular relevance lies in its focus on multimedia learning and its evidence-based principles for analysing instructional video design.

1.5.4 Psychology

Psychological and educational-psychological theories provide models for interpreting the learner-side processes examined in the thesis. Control-Value Theory conceptualises achievement emotions as related to learners' appraisals of control and value (Pekrun 2006), while the Meaning and Attention Components model explains boredom in relation to perceived meaning and the fit between attentional demands and available resources (Westgate and Wilson 2018). These theories are compatible with the study's epistemic

orientation because they assign explanatory importance to the learner's appraisal and interpretation of a situation without requiring the adoption of a strong constructivist ontology.

The thesis also draws on categorical and dimensional approaches to emotion. Basic Emotion Theory contributes concepts used in the classification of facial movements and therefore has an important methodological role in the study. It is not, however, adopted as the exclusive ontological account of emotion. Contemporary emotion research includes substantial disagreement concerning the extent to which discrete emotions possess unique and universal physiological or expressive signatures (Ekman 1992; Barrett 2017; Barrett et al. 2019). The empirical analysis, therefore, distinguishes between software-generated emotion labels and participants' reported emotional experiences.

1.5.5 Affective Computing

Affective computing contributes primarily to the measurement architecture of the thesis. The principles, goals and strategies shared by the AAA (Advanced, Analytic and Automated) model of affective computing (D'Mello, Dieterle and Duckworth 2017) provided important guidance for the present research, particularly in highlighting the potential of complementing self-report measures with additional measurement approaches to advance the assessment of academic emotions. The value of affective computing in the present study lies primarily in identifying temporal changes, peaks, recurring facial patterns, and associations with instructional features. Cross-method comparison then provides evidence concerning the extent to which these patterns correspond with learners' own accounts.

1.6 The structure of the thesis

The thesis is composed of three main sections. The first part, comprising Chapters 1 to 6, provides the theoretical background for the study by reviewing the main concepts, definitions, and previous research findings relevant to the dissertation's focus. In the course of the literature study, the dissertation's focus was progressively refined by identifying thematically relevant areas and selecting subtopics that address key research gaps and conceptual puzzles within the field (see Figure 2).

The second part provides the methodological foundation for the empirical study and consists of two closely connected chapters. Chapter 7 serves as a transition between the theoretical review and the empirical methodology by synthesising the literature on methods for measuring academic emotions, with particular attention to automated facial expression analysis. Chapter 8 then presents the study's research design and methods in the more conventional form of a methodological chapter. This extended methodological section is justified both by the methodological focus of one of the research questions and by the use of facial analysis and related analytical procedures, which remain relatively uncommon in instructional communication research. Together, the two chapters move from a review of available methodological approaches to the specific design adopted in the present study.



Figure 2: Delimitation of dissertation topics based on relevance and research gaps

Note. Author's own compilation

2 STUDENT ENGAGEMENT

The concept of student engagement has developed across an expanding range of disciplines and perspectives, generating rich debate (Reschly and Christenson 2012) while also becoming more nuanced and multifaceted. The current chapter provides an overview of this development, including the definitions that have shaped the field, the changing emphases across research traditions, and the components of engagement examined in the present thesis.

2.1 Definitions

Student engagement, due to its meta-construct nature, lacks a universally accepted definition (Kahu 2024). The present section will therefore enumerate several definitions considered relevant to the current research and formulate a working definition for the current thesis.

Kahu and Nelson defined engagement as "an individual student's psychosocial state: their behavioural, emotional and cognitive connection to their learning" (Kahu and Nelson 2018, p. 59), mediated by self-efficacy, emotions, belonging, and wellbeing. They, however, place greater emphasis on the role of external, meso-, and macro-level factors, underscoring that student engagement is strongly influenced by socio-ecological factors, including classroom, institutional, community, and political influences, which are shaped by the broader cultural and political environment.

Approaching engagement at the *individual level* to understand its function and detect it required a definition that supports precise scientific measurement. D'Mello et al. (2017, p. 106) defined engagement as 'a goal-directed state of active and focused involvement in a learning activity'. This definition was refined by Dubbaka and Gopalan (2020, p. 447), who conceptualised it as 'a measure of a learner's attention (arousal) combined with their level of positive/negative emotions (valence)'. The current dissertation adopts D'Mello's concept, while referring to the elements of the above definitions, defining the concept as follows:

> *Student engagement is a dynamic and context-dependent state of goal-directed, active and focused involvement in learning, expressed through interrelated behavioural, cognitive and emotional dimensions.*

2.2 The evolution of the concept

The concept of student engagement has become a central subject of research, policy, and pedagogical practice only in the past 30 years. Its importance is recognised as crucial from the perspectives of social equity (Asmar, Page and Radloff 2015; O'Shea and Delahunty 2018; Trowler 2015), institutional effectiveness (Christenson, Reschly and Wylie 2012), return on public investment in higher education and students' academic success and personal growth (Kahu 2013).

The initial attempts to map the factors influencing learning effectiveness were mainly focused on behavioural elements associated with successful learning and disengagement (Carroll 1963), but research soon revealed that student engagement *is more than a way of behaving*: it is also a state of mind that affects emotions, cognitive processes and attitudes. Changing student behaviour or increasing time allocated for learning will not necessarily increase student engagement (Reschly and Christenson 2006).

The first conceptual framework that defines and explains student engagement was developed by Fredricks, Blumenfeld, and Paris (2004), who, applying a socio-psychological approach, distinguished three main dimensions of the concept: *Behavioural engagement* refers to students' participation, time and effort investment into their learning, and to the lack or presence of unconstructive behaviours. *Emotional engagement* refers to students' emotional responses, which shape their emotional responses towards the instructor, the academic institution, their peers, and, consequently, to learning. *Cognitive engagement* can manifest in the quality of learning and in the application of strategies that support deep and self-regulated learning. This three-dimensional model has been expanded to include additional dimensions, such as agentic engagement, which refers to learners' proactive contributions to the learning process (Reeve and Tseng 2011; D'Mello, Dieterle and Duckworth 2017). Further dimensions proposed for the original three-dimensional model included social and collaborative dimensions, reflecting the development of relationships and learning-related partnerships, respectively (Redmond et al. 2018).

Later theories, representing a psychological perspective, increased their focus on emotions and intrinsic motivation, arguing that positive emotions such as pleasure and interest related to learning result in a 'substantive engagement' (Kahu 2013, p. 762). This approach has received increasing prominence in the relevant literature during the past decades, underscoring that emotions like enjoyment, boredom, or anxiety not only profoundly impact attention, motivation, strategy use, and ultimately academic success but also show a reciprocal causation between engagement and emotions: emotions impact students' behavioural, motivational engagement and their achievement, and these all shape students' emotions (Pekrun and Linnenbrink-Garcia 2022).

The global pandemic had a radical effect on the approach to student engagement, bringing a relational turn in education. Changes have highlighted the importance of inclusion and diversity in research on engagement, as well as the crucial role of interactions in engagement, whether between peers or between students and lecturers. (Matthews et al., 2024, see also Section 5.2, Instructional communication in the context of online learning). Moreover, the power of "relationship-rich education" (Felten and Lambert, 2020, cited by Kift, 2024) and the substantial role of *belonging* received heightened attention due to the experiences of this period.

The holistic perspective reframes student engagement as a dynamic process deeply embedded in the socio-cultural context, shaped not only by individual student characteristics but also by institutional structures, psychosocial influences, and their interaction. Building on the statements of the three above-discussed perspectives, Kahu and Nelson (2018) proposed a conceptual framework demonstrating the dynamic relationship between the components and the effects of student engagement within the educational ecosystem, viewing engagement as a process occurring at the intersection of the individual, their background, and the institution's practices. This model embodied the macro-, meso-, and micro-level mechanisms that mediate students' psychological engagement.

The *macro* level factors include wider socio-political, policy and equity conditions that shape participation; the *meso* level conditions refer to institutional cultures, curricula and measures; and the *micro* level is represented by the "educational interface" (Kahu and Nelson 2018) or "engagement interface" (Trowler et al. 2022) where students' backgrounds, dispositions and day-to-day learning interactions meet. While macro and

meso-level perspectives examine broader institutional forces and conditions that structure engagement, research focusing on *micro* level processes investigates psychosocial mechanisms and interactions that activate or weaken engagement. In recent years, factors influencing engagement at the *micro* level have attracted the *greatest research attention,* which also serves as the focus of the *present thesis*. This is presumably because, at this level, factors are the most alterable and interventions can be tested and implemented, regardless of the conditions established at the meso- and macro levels.

2.3 Conceptualisation

While the behavioural, psychological, socio-cultural and holistic approaches discussed in Section 2.2 provide broader *theoretical* perspectives for interpreting student engagement, the current chapter presents several *facets* (also mentioned as dimensions) of engagement, referring to the multiple dimensions through which engagement is expressed. The first conceptual framework that defines and explains student engagement was developed by Fredricks, Blumenfeld, and Paris (2004), who, applying a socio-psychological approach, distinguished three main dimensions of the concept: *Behavioural* engagement refers to students' participation, time and effort investment into their learning, and to the lack or presence of unconstructive behaviours. *Emotional* engagement refers to students' emotional responses, which shape their emotional responses towards the instructor, the academic institution, their peers, and, consequently, to learning. *Cognitive* engagement refers to the cognitive quality of learning and the application of strategies supporting deep and self-regulated learning.

The variations in the facets proposed by the studies cited above delineate a tendency that confirms the validity of the original three-facet model of (behavioural, cognitive and emotional) engagement, while also noting that it does not adequately capture all aspects of student engagement in higher education, and emphasising the importance of the broader context encompassing social, collaborative and socio-cultural elements. The current thesis aims to investigate the role, influencing factors, triggers and measurement opportunities of *emotional engagement*.

2.3.1 Distinguishing student engagement from motivation

In everyday discourse, the terms "engagement" and "motivation" are often used synonymously. Most theoretical analyses of these concepts agree that the two concepts should be distinguished. However, they offer different, distinctive aspects, and the relationship between the two concepts is viewed differently. Following a lengthy debate, certain aspects of these differences became widely accepted (see Appendix A). The most important differences between the two concepts stem from their origins: Motivational theories have a longer history and are rooted in cognitive, developmental, and social psychology. Research on engagement emerged within educational science, partly from efforts to better operationalise and observe motivation in educational contexts, with a focus on practical implications for interventions (Skinner and Raine 2022). Motivation and engagement, therefore, can be interpreted as two different types of concept: while student engagement is a broad construct, encompassing the area of motivation as well, motivation also ‘exists as an independent and worthy area of study’ (Reschly and Christenson 2022, p. 19), discussed by a range of theories.

2.3.2 Is disengagement the opposite of engagement?

A question debated for several years was whether engagement and disengagement were the opposite ends of the same continuum or distinct constructs arising from distinct variables. Since then, numerous findings have confirmed that disengagement is not synonymous with low engagement but is a distinct construct with separate *emotional* and *behavioural* components (Reschly and Christenson 2022; Wang, Degol and Henry 2019). This approach enables the two-dimensional analysis of disengagement, specifically in relation to engagement, recognising that high levels of engagement and high levels of disengagement may be experienced at the same time, either due to overchallenge (Salmela-Aro, Tang and Upadyaya 2022) or due to the use of inadequate learning strategies (Gobert et al., 2015, as cited by D’Mello, Dieterle and Duckworth 2017).

2.4 Focus of the thesis related to student engagement

The above description reveals that engagement is a complex construct interrelated to several elements of the educational ecosystem: the level and the conditions for student engagement depend on the national educational policies, the institutions’ internal

strategies, the social-economic and cultural background of the learners, the physical conditions of the learning environment, learners' personality and the instructors' personal and professional competences. The current thesis examines engagement at the micro level, focusing on relationships between instructional communication and engagement.

Applying a psychological approach to student engagement, the inquiry's primary aim was to examine emotional engagement; however, given its close association with cognitive engagement, the relationship between the two facets was also examined. The research questions and theoretical foundations of the present thesis clearly align with the *holistic, relational focus* articulated by Kahu & Nelson (2018) and Kift (2024), which considers the emotional, contextual, and socio-cultural dimensions of student engagement in higher education. The interpretation of the research findings takes into account the realisation that SE is not a concept that can be understood within the framework of a static and linear system; rather, it changes dynamically over time and may even exhibit contradictions or variations in the levels of different types of engagement at any given moment (Trowler et al. 2022).

Finally: Even though the current study approaches engagement by focusing on the potential of instructors as actors, it intends to draw attention to the words of Ella Kahu, the developer of the first and since widely adapted conceptual framework for student engagement, who warns that 'there are many ways to improve student engagement, and the responsibility for this lies with all parties: students, teachers, institutions, and governments' (Kahu 2013, p. 769).

3 ACADEMIC EMOTIONS

Emotions play an important role in student engagement by influencing its cognitive, motivational, and behavioural dimensions, while the reciprocal relationship between emotions and engagement may further reinforce these effects (Pekrun & Linnenbrink-Garcia, 2022, pp. 125–126).

Figure 3 provides a simplified representation of this process: The learning environment (including students' self-regulation, goal structure, motivation, and the strategies they use) shapes their appraisals of the situation and their perceived control and competence, which in turn result in several types of emotions. The types, valence, intensity, and duration of these emotions will affect the levels of student engagement and disengagement. The level of cognitive, emotional and behavioural engagement has a direct effect on student achievement. Students' performance and results have an influence on the learning environment:

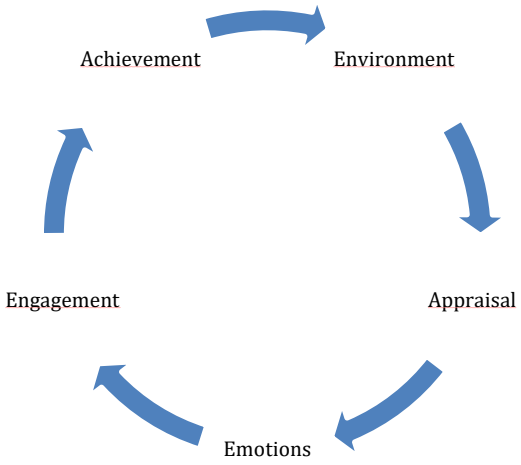


Figure 3: Reciprocal causation of academic emotions, engagement, and their antecedents and outcomes. Author's own compilation based on Pekrun & Linnenbrink-Garcia, 2022.

This is why the current thesis strives to examine, measure, and draw conclusions about engagement – a complex, multifaceted, intangible, and not directly observable construct – based on students' emotions experienced during learning.

3.1 Emotions

According to their most widely applied definition, *emotions* are psychological responses given to an external or internal stimulus that is perceived as important or relevant for the individual (Scherer 2005). Emotional processes usually comprise five components: (1) a cognitive element referring to the cognitive evaluation of the event, (2) a neurophysiological response, i.e., physical feelings, (3) a motivational aspect, (4) a behavioural-expressive process, e.g., body posture and facial expression and (5) an affective component, including a subjective emotional experience. The current thesis builds on the definition proposed by Pekrun & Goetz (2024, p. 75), treating emotions as responses to events and to "ongoing situations, actions, persons, and various other phenomena", collectively termed "objects".

Within the system of affective phenomena, *emotion* has been defined and differentiated from *mood*, *emotional traits* (Davidson et al. 1994), attitude, feelings and affective experiences (Scherer, 2005) based on their duration, antecedents and intensity. Emotion is a transient, easily changing, but intense episode that usually lasts seconds/minutes and is triggered by a given situation or object. *Feelings* are conscious, subjective interpretations of emotional states that may persist longer than the original emotion and change more gradually than emotion (Scherer, 2005). *Mood* is a longer duration (hours or days) but less intensive state, often lacking a clearly identifiable trigger, indicating a background affective state that influences perception and behaviour over time (Csépe, Győri and Ragó 2007). Emotional *traits* are stable dispositions that influence how individuals generally feel or behave across situations. *Affective experiences* (e.g., preferences, attitudes, and affective styles) constitute a broader category, encompassing any relatively stable subjective experience with positive or negative valence.

The main function of emotions is to evaluate changes in the environment: they inform the individual about the value, urgency, and importance of an event, thus helping them pursue the appropriate behaviour or interaction in the given situation (Scherer 2005).

3.2 Emotional taxonomies

Emotional states can be distinguished by their quantitative features (intensity) or by their qualitative characteristics (classification into a particular emotion category). Several taxonomies attempt to define, compare and categorise emotions in line with these features.

3.2.1 Categorical taxonomies

The *categorical taxonomy approaches emotions as distinct categories and compares* them according to their affective, cognitive, motivational, physiological, and expressive components (Ekman 2003; Plutchik 1980; Scherer and Moors 2019). To identify emotions, they use categories provided by the language's semantic system, based on semantic maps that represent the typical patterns of their occurrence and components (Lazarus 1991; Scherer 2005). However, the cultural context plays a fundamental role in how emotions are experienced, expressed, and labelled, and, consequently, the use of these categories and semantic maps reveals cultural differences.

This was the main reason for Paul Ekman's (1975) attempt to identify the emotions and emotional states that are universally expressed and consistently interpreted, regardless of culture, geographic location, and social background. The hypothesis behind the *Basic Emotions Theory* was tested by presenting photographs of facial expressions to people from different cultural backgrounds and asking them to choose the emotion that best matched each expression based on a short story. Based on a series of experiments and observations, he identified 7 emotions that were deemed to

- show a degree of cross-cultural similarity in their expression and recognition,
- that might be innate,
- that show a well-distinguishable psychological pattern with characteristic thoughts and subjective experience.
- These expressions have their origins in the behaviour of other mammals in situations similar to those faced by humans (e.g. when signalling hostile or cooperative attitudes), as they might have had important adaptive functions.

The Basic Emotions Theory identifies 7 basic emotions, in line with the above aspects: happiness, sadness, fear, anger, surprise, disgust, and contempt (Ekman and Cordaro 2011; Ekman and Friesen 1975). Applying a similar methodology, this research was later

expanded by several subsequent studies to explore a broader range of emotions and their most frequent physiological expressions. Keltner and Cordaro (2017) identified 24 emotions across 10 cultures that can be considered universal, and they additionally refined the methods used in their examinations: Instead of relying on static images and collecting only visual data, they used multimodal data, including voice, body posture, touch, and even music. In addition to recognising certain emotions, they also considered the level of recognisability, culturally influenced patterns of expression within a given culture, and physiological changes associated with the given emotional state. Participants in their research were asked to provide free responses rather than choose from a limited set of emotions when encoding emotions.

Basic Emotion Theory is relevant to the present study primarily because categories derived from this tradition inform widely used systems for coding and automatically classifying facial movements. Its use in the methodology should not, however, be interpreted as an assumption that each discrete facial configuration provides a direct or context-independent representation of an internally experienced emotion. The status and universality of discrete emotion categories remain debated. Constructed-emotion accounts, for example, conceptualise emotional episodes as context-sensitive constructions involving prior experience, conceptual knowledge, bodily information, and situational interpretation (Barrett 2017). A broad review of facial expression research likewise found substantial variation in the relationship between facial movements and emotion categories. It cautioned against inferring specific internal emotional states directly from isolated facial configurations (Barrett et al. 2019).

Basic Emotion Theory also provided the basis for the Facial Action Coding System (FACS), in which observable facial movements are described as Action Units; its relevance to automated facial analysis is discussed in Section 7.5. In the empirical chapters of this thesis, categorical FaceReader outputs are therefore treated as model-generated indicators of facial activity whose psychological meaning requires contextual and cross-method interpretation.

3.2.2 Dimensional taxonomies

The dimensional taxonomies classify emotions based on continuous dimensions such as pleasantness or activation. They draw attention to the significance of gradations and blends between emotions, allowing for more flexible representation of affective states.

One-dimensional models categorise emotions based on a single dimension, such as positive and negative affectivity. These models use a variable called *valence* to describe emotions along a single scale ranging from pleasant (high valence) to unpleasant (low valence) (Watson and Tellegen 1985a). The current study applied the Positive-Negative Affectivity Scale developed based on this model in the self-report questionnaire (see Section 8.4.1). Russell's *circumplex model of affect* (Russell, 1980, see Figure 4) expands this model with the dimension of activation, capturing a more refined emotional complexity, allowing for more nuanced categorisation. This model organises emotions in a two-dimensional space, along the *physiological activation* (arousal) and *pleasantness* (valence) dimensions. Emotions can thus be organised into four categories: positive-activating (e.g., excitement), positive-deactivating (e.g., relaxation), negative-activating (e.g., anger), and negative-deactivating (e.g., sadness).

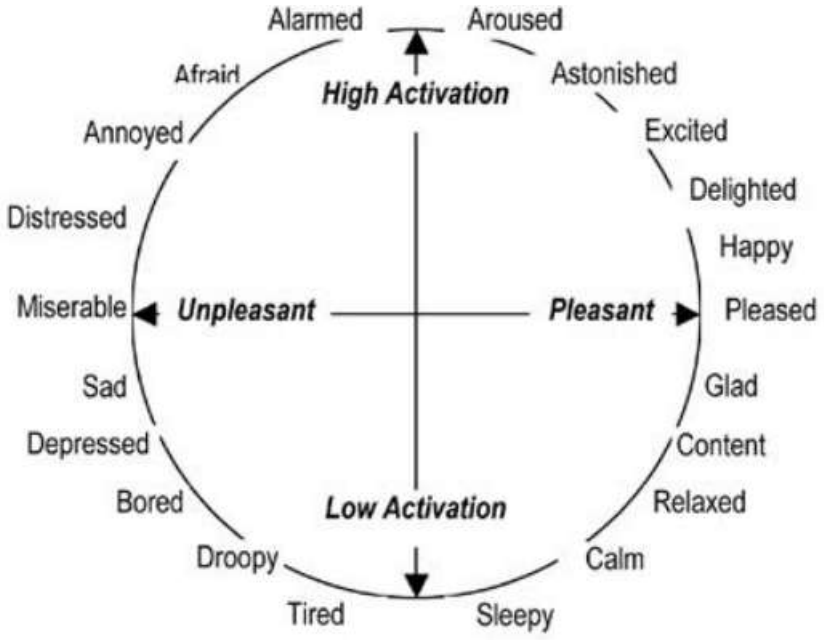


Figure 4: The circumplex model of affect developed by Russell.

Source: Han and Patterson 2007, p. 333

The research presented in this thesis drew on this framework because it is widely used in related research, as reflected in FaceReader's emotion categorisation system (see Section 7.5).

Two- and three-dimensional taxonomies of emotions add further dimensions to the above-discussed ones, such as temporality, indicating the emotion's appearance before, during, or after a particular process or outcome (Pekrun, Marsh, Elliot, et al. 2023) or the level of the perceived control over the situation (Bradley and Lang 1994).

3.3 The role of academic emotions

Education has traditionally been considered a field in which rationality is more significant than emotion; moreover, several approaches have suggested that emotions may endanger the learning process and distract students' attention (Dewey 1944; Seneca 60). Over the last three decades, a growing number of studies have shown that emotions play a decisive role in learning, and their effects on learning have become increasingly evident and widely investigated. The first learning-related emotion examined in depth was exam anxiety, which showed a strong correlation with achievement level. The influence of certain emotions on learning engagement has become one of the areas of educational theory and psychology that has attracted interest and enquiry in the recent two decades (Pekrun, 2006; Titsworth et al., 2013). The current section summarises their findings and their effect on learning (see the references below), given that the present thesis aims to contribute new results to these findings.

3.3.1 The Control-Value Theory

The two approaches - dimensional and categorical - discussed above are often seen as complementary rather than mutually exclusive in contemporary emotion research. The Control-Value Theory of Achievement Emotions (CVT, Pekrun, 2006) is the main appraisal based explanatory theory of achievement emotions. It is rooted in categorical emotion theory, given its focus on discrete emotions and their specific antecedents and functions, but it has been expanded and enriched with dimensional components, particularly in its later developments, and can thus be described as a categorical foundation with dimensional integration.

The CVT integrates key motivational traditions by integrating expectancy-value theories, Bandura's social-cognitive approach, and Weiner's attribution theory.

- According to expectancy-value theories (Atkinson 1957; Heckhausen 1977), motivation depends on students' *expectations* and the *importance* they attribute to the activity or its outcomes.
- The concept of self-efficacy, as defined by Bandura, serves as a fundamental basis for perceived control, referring to students' *beliefs* regarding their *ability* to complete learning tasks successfully (Bandura 1977).
- Weiner's attributional theory proposes that students' emotional reactions depend on how they explain success and failure, especially whether they attribute outcomes to themselves or to external factors, and whether they see these causes as controllable or uncontrollable (Weiner 1985).

The CVT model builds on these traditions, identifying control and value as the two central factors in understanding students' motivation and emotions. Perceived *control* is therefore defined as the extent to which students believe they can influence learning activities and outcomes. *Value*, on the other hand, refers to the importance, meaningfulness, interest, or cost of these activities and outcomes. Students are less likely to invest effort, persist, or regulate their learning effectively if they do not believe they can affect success, or if success has little personal significance. Control and value are therefore highlighted in the theory because, together, they help explain why students engage, persist, withdraw, and respond emotionally in learning situations.

The Three-Dimensional Taxonomy of Achievement Emotions is an expanded version of the CVT (2023), distinguishing emotional states according to the following criteria (see **Hiba! A hivatkozási forrás nem található.**): Are they positive or negative? (pleasantness/valence, see Section 3.2.2)

(a) Do they activate the physiological system, increase the energy level and have a mobilising effect? (arousal)
(b) Are these emotions connected to the learning *process* (*activity* focus) or to the *outcome* of learning? In the latter case, do they refer to a past event (*outcome-retrospective*) or to an expected, future outcome (*outcome-prospective*)?

Table 1: The three-dimensional taxonomy of achievement emotions

	Positive		Negative	
Object focus	Activating	Deactivating	Activating	Deactivating
Activity	**Enjoyment** Excitement	**Relaxation**	**Anger** Frustration	**Boredom**
Outcome–prospective	**Hope** Anticipatory joy	**Assurance**	**Anxiety**	**Hopelessness**
Outcome–retrospective	**Pride** Retrospective joy Gratitude	**Relief** Contentment	**Shame/Guilt** Anger	**Disappointment** Sadness

Source: Pekrun, Marsh, Elliot, et al. 2023, p. 149

A recent large-scale study (Pekrun, Marsh, Elliot *et al.* 2023) examined the empirical robustness of the above taxonomy and, additionally, explored the relationships between these emotions and variables indicating learning engagement and effectiveness, such as perceived control, intrinsic value, enthusiasm, effort and performance. The results supported the validity of the taxonomy's 3 main aspects and provided robust empirical evidence for the relationships between these emotions and features of effective and engaged learning.

As a general tendency, it can be stated that

- *positive and high-arousal emotions* show the strongest association with the behaviours and cognitive processes that contribute to effective and engaged learning.
- The opposite end of the spectrum, i.e., *negative and deactivating* emotions, seems to be the most detrimental to effective learning.
- *Positive emotions with low arousal levels and negative emotions with high arousal* show a more complex picture. Their effects can sometimes be beneficial, but to a limited extent.

Table 2 follows the categorisation of the Three-Dimensional Taxonomy of Achievement Emotions (Pekrun, Marsh, Elliot, et al. 2023), grounded on the Control-Value Theory of achievement Emotions (CVT, Pekrun, 2006), according to their effect on learning effectiveness, i.e. whether they support, hinder or have a varying relationship with learning engagement, depending on situation and duration. It includes an expanded list of emotions reported in empirical research on academic emotions (see the references below Table 2).

It is important to note that although the emotions listed in the Three-Dimensional Taxonomy (Table 1) were based solely on self-report questionnaire results, several of the studies cited below used additional measurement methods.

Table 2: The effect of emotions on academic engagement, motivation and performance (see the sources below the table)

Object focus (time factor)		**Predominantly positive associations with engagement and learning**	**Predominantly negative associations**	**Mixed or context-dependent associations**
Outcome-prospective (future)	*High arousal emotions*	hope assurance curiosity surprise excitement enthusiasm	panic	anxiety
	Low arousal emotions		hopelessness	
Activity related (present)	*High arousal emotions*	enjoyment / happiness flow excitement	nervousness frustration	(anger) confusion
	Low arousal emotions	relaxation	boredom tiredness feeling of isolation sleepiness depression	relief neutral state sadness
Outcome-retrospective (past)	*High arousal emotions*	pride gratitude	shame guilt disgust	
	Low arousal emotions	contentment	disappointment	

Author's own compilation, based on Almohammadi et al 2017; Altuwairqi, 2021; Bian et al, 2018; Boekaerts, 2007; Bosc, 2015; Dhall et al, 2018; D'Mello et al, 2007; D'Mello and 2009, 2017; Dragon et al 2008; Grafsgaard et al, 2013; Ketonen 2017; Kim and Hodges, 2012; Mottet et al., 2006; Pekrun et al., 2011; Wingenbach, 2018; Whitehill et al, 2014.

3.3.2 Emotions predominantly positively associated with engagement and learning

Every high-arousal emotion, including positive ones (e.g., pride or hope), may distract students' attention from the task as it consumes cognitive resources and reduces task focus (Pekrun 2006, p. 200; Terzis, Moridis and Economides 2013). Enjoyment and flow related to the learning activity are exceptions, as they foster focus and support cognitive work. Nevertheless, in general, positive achievement emotions support not only psychological wellbeing and life satisfaction but are also related to greater activity and achievement (Fredrickson 1998; Williams, Childers and Kemp 2013). They strengthen intrinsic and extrinsic motivation, facilitate the application of flexible and creative learning strategies, and foster the adaptation of learning to the actual goals, thus supporting self-regulated learning (Pekrun 2006). Positive and activating emotions show positive correlations with attention and the use of flexible learning strategies, promoting elaboration and critical thinking, effort and self-regulation while being negatively related to task-irrelevant thinking. Moreover, the relationship between enjoyment and the use of flexible learning strategies shows reciprocal causation (Pekrun, Marsh, Elliot, et al. 2023).

Both deactivating (i.e., low-arousal) negative emotions, such as boredom, hopelessness, or disappointment, and high-arousal negative emotions (such as anger, anxiety, and shame) showed negative correlations with the use of flexible learning strategies and effort, and were positively related to irrelevant thinking. These effects are noticeably detrimental to motivation and interest, undermine the use of cognitive learning strategies such as cognitive elaboration, and thus lead to superficial information processing (Chiu et al. 2014; Goetz et al. 2024; Pekrun 2006).

3.3.3 Emotions with mixed or context-dependent associations with engagement and learning

The effects of some positive emotions with a low arousal level and some negative emotions with high arousal show a complex picture. Their effects can sometimes be beneficial or disruptive, depending on their intensity, duration, and students' perceived self-efficacy.

- *Anger*: Self-report-based research clearly shows a negative relationship between *anger* and performance, perceived intrinsic value and self-regulation.

- *Confusion*: Several empirical studies concluded that frustration could be beneficial to learning, but only under some circumstances: temporal occurrence of confusion and frustration can be a part of experiencing interest (Hidi & Harackiewicz, 2000). This is also consistent with the distinction between interest and enjoyment. Westgate and Wilson (2018) note that some conditions that increase interest, such as novelty and complexity, may reduce enjoyment, suggesting that these states are not interchangeable and may sometimes show partly opposing affective patterns. Interest may therefore involve cognitive effort and need not correspond to an immediately pleasant experience. When confusion results from a cognitive conflict intentionally included in the learning process (especially with complex problems and topics), and when the student receives the support needed to handle this conflict or when instructional help is available, confusion can effectively promote deep learning. If confusion persists beyond a certain duration, confusion leads to frustration, boredom, and disengagement (D'Mello et al., 2014); thus, the key to applying this strategy is to define the optimal duration and intensity of confusion in a given learning situation (Arguel et al. 2019).
- *Anxiety and shame*, as negative activating emotions; they tend to be associated with more task-irrelevant thinking, lower interest and intrinsic motivation, reduced use of flexible learning strategies, and weaker self-regulation; therefore, they are generally regarded as emotions that tend to impair learning effectiveness. At the same time, because they are closely linked to failure avoidance, they can also foster avoidance-oriented extrinsic motivation, increase effort investment, and promote more rigid strategies such as rehearsal (Pekrun, Marsh, Elliot, et al. 2023). Nevertheless, anxiety shows a curvilinear relationship with motivation: when failure is perceived as possible but not inevitable, anxiety may mobilise effort and trigger avoidance-oriented motivation (Pekrun 2019, 2024).
- *Relief* is defined by Pekrun as "the non-occurrence of anticipated failure" (Pekrun 2006, p. 323) or a result of a positive emotional state resulting from withdrawing from a challenge or task (Carver and Scheier 2014).
- *Sadness*: Though sadness is primarily classified as a deactivating and low-valence emotion (Pekrun 2006), contrary to relief, approach behaviours towards higher levels of self-regulated learning are sometimes accompanied by sadness resulting

from doubts about goal attainment; sadness can also be experienced after failure or withdrawal. (Carver and Scheier 2014).

- A study distinguishing emotions along the dimensions of motivational direction (approach vs. avoidance) and emotional valence (positive vs. negative) found that negative valence emotions *such as anger*, *frustration*, or even *hostility*, *irritability* and *depression* can be associated with approach motivation, i.e., with behaviours reflecting a goal-directed approach, activated in response to difficulties or setbacks, aiming to take action and resolve the situation (Carver and Harmon-Jones 2009).

The majority of the emotions discussed in this section fall into the unpleasant category. The reason for their ambiguous relationship with engagement is partly due to their duration: while the temporary experience of negative emotions is an integral part of the learning process, prolonged persistence has the opposite effect. In the case of the other emotions listed above, facial expressions do not necessarily indicate negative emotions; rather, they reflect cognitive processes that, at the level of facial expression, are accompanied by movements similar to those associated with sadness and anger.

4 BOREDOM

This chapter aims to present the diverse perspectives offered by theoretical approaches to boredom. Particular attention is given to the numerous open questions for which research findings remain inconclusive or contradictory. The thesis aims to contribute to these ongoing debates by using its empirical findings to clarify these unresolved questions and support a more precise understanding of the issues involved.

4.1 A growing field of research

From ancient times, literature and philosophers have often described and addressed boredom, but boredom experienced in the context of learning – academic boredom - has received highlighted attention and emerged *as a field of research* only in the past 2-3 decades. This attention has grown exponentially since then (see Figure 5).

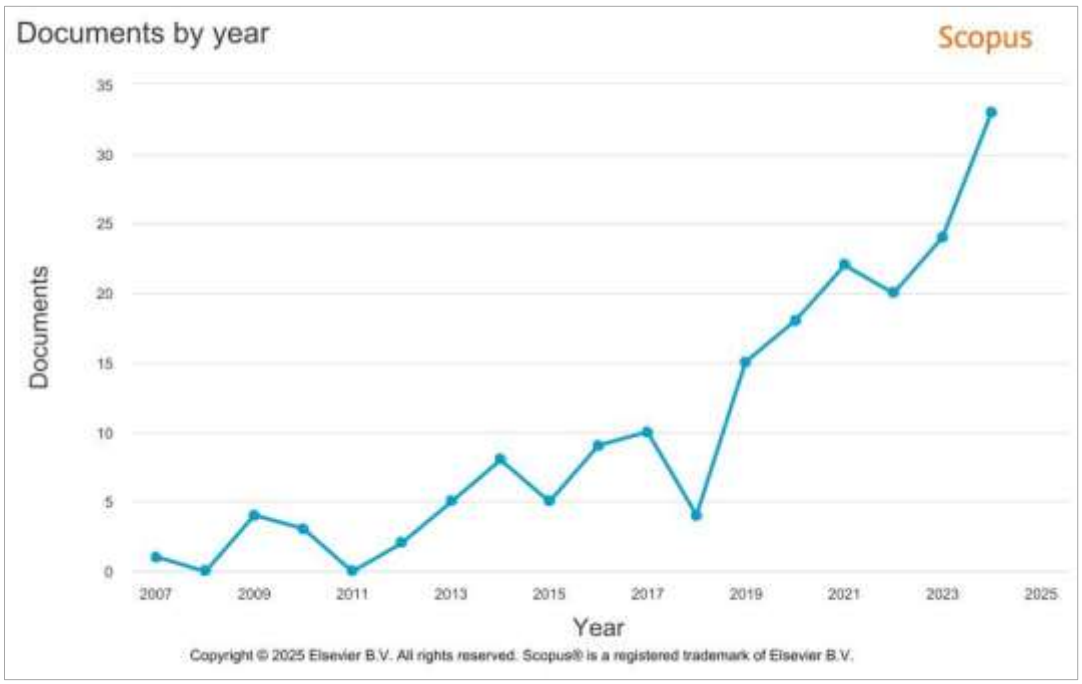


Figure 5: Annual number of publications on academic boredom indexed in the Scopus database

While between 1997 and 2013, boredom was the 4th most frequently studied academic emotion in the context of affective computing (Wu, Huang and Hwang 2016), between

2010-2017, from all the emotional states examined, boredom was moved to the *top* of the frequency analysis list of emotions (see Figure 9, Yadegaridehkordi *et al.* 2019). This tendency has accelerated during and since COVID: the rate of research articles discussing *the topic of boredom was* 30% higher in 2020 than in 2019, making it one of the fastest-growing topics and drawing a lot of public attention, as shown by global media coverage during the pandemic (Neuroskeptic, 2020, as cited by Bieleke, Wolff and Martarelli 2024.

Across studies on boredom in general or in different areas of life, the proportion of articles focusing on academic boredom has increased significantly over the past 12 years, from 1.5% to 2.4% (Goetz et al. 2024). The growing diversity and intensity of academic boredom research can also be interpreted as a strong signal of its broad relevance while complicating efforts to maintain a comprehensive overview and identify commonalities across disciplines. The *Routledge International Handbook of Boredom* was a pioneering endeavour that brought together eminent scholars to present their research and perspectives on this phenomenon.

4.2 Conceptualisation and theoretical accounts of boredom

This section synthesises the key findings of theoretical and empirical research on academic boredom conducted over the past two decades. Despite being a common experience, boredom remains conceptually and theoretically contested, with differing definitions, assumptions, and sometimes contradictory findings across the major explanatory frameworks and empirical studies.

4.2.1 Definitions

Definitions of boredom partly reflect broader differences in its conceptualisation. Although it is commonly treated as an emotion, it has also been conceptualised as an affective, cognitive-affective, or motivational state (see Section 4.3.1).

By one of the most widely used definitions, boredom is an "aversive state of wanting but being unable to engage in satisfying activity" (Eastwood et al., 2012, 482). A recent definition (Wolff et al., 2024, pp. 8-9.) argues that the effects and antecedents of boredom can be so different, depending on the context, that a more universal definition can be found by focusing on its original, evolutionary function: it thus can defined as "a state of insufficient function utilisation, in which reward prediction error has been minimized"

(Wolff, Radtke and Martarelli 2024, p. 22), with other words, an unpleasant state signalling that the individuals' functions are not adequately utilised and in which nothing unexpected or meaningful happens. The function of this unpleasant feeling is to trigger an "undirected exploration" motivating a better utilisation of their capacity (Wolff, Radtke and Martarelli 2024, p. 12).

The Meaning and Attentional Components Model (MAC) of boredom (Westgate and Wilson 2018) defines boredom as a state in which the level of *attention* and the perceived meaningfulness of the task are not aligned with the needs of the activity. This usually means that the individual experiencing boredom *lacks attention* or a sense of *meaning*. Extending this account, Westgate and Steidle (2020) conceptualise boredom as an informative signal of unsuccessful meaningful engagement that motivates change, either by restoring attention or meaning within the current activity or by shifting towards a more rewarding alternative.

The Control-Value Theory conceptualises boredom as a coordinated emotional response to activities perceived as lacking value or meaning, involving unpleasant affect, attention-related cognitive processes, motivation to escape, physiological deactivation, and characteristic expression (Pekrun and Goetz 2024).

Finkielsztein (2024) notes that recent definitions have emerged from both psychology (Eastwood et al. 2012; Elpidorou 2024; Pekrun and Goetz 2024; Westgate and Wilson 2018) and from philosophy (Van Tilburg and Igou 2012). His recent definition synthesises the main elements of these disciplinary approaches developed over the past 10–15 years, describing boredom as "a transient, negatively perceived, transitional emotion or feeling of listless and restless inattention to and engagement withdrawal from interacting with one's social and/or physical environment caused distinctively by an atrophy of personally-valued meaning, the frustrated need for meaning" (Finkielsztein 2024, p. 14).

4.2.2 State and trait boredom

As an individual experience, boredom can be examined from two perspectives (Farmer and Sundberg 1986).

(a) The extent to which individuals tend to experience boredom shows considerable variation as it is strongly influenced by personality characteristics. This tendency is referred to as *trait boredom*.
(b) In contrast, *state boredom* refers to the momentary experience of boredom that varies depending on context. While the intensity and frequency of state boredom are also highly individual, identifying its patterns and context-specific antecedents can help reveal the mechanisms that trigger boredom.

Trait and state boredom raise fundamentally different research questions and exhibit distinct patterns when examined across various factors, stimuli, and each other. As this thesis aims to explore such patterns and relationships, it focuses solely on the features and precursors of *state boredom*. This focus is further supported by experience-sampling evidence suggesting that much of the variation in everyday boredom is attributable to situational rather than stable individual differences, which also calls for caution when drawing causal conclusions from trait-boredom associations (Westgate and Steidle 2020).

4.2.3 Academic boredom

Boredom is characterised as "relatively universal" by Pekrun and Goetz (2024), meaning that, according to the low number of comparative studies, the proximal antecedents, consequences, the emotions accompanying boredom, and the coping and preventing strategies managing it, seem to be similar, independently of cultures, contexts, and individual differences. As regards its intensity and process, it can show differences depending on the context of boredom. Consequently, it is worth studying the specific features of boredom across different situational domains, e.g., education, sport, the workplace, social interactions, food consumption, or leisure activities. Topic boredom is a content-specific occurrence evoked by a specific content or learning activity (Bieleke, Wolff and Martarelli 2024b). The present thesis focuses on boredom experienced in the context of education, called *academic boredom*, focusing on its features within higher education.

Academic boredom can be defined simply as "boredom experienced in academic settings" (T. Goetz et al. 2024, p. 226) or, in more detail, as "an achievement-related emotion experienced by students in the context of university education (mainly during class- or

learning-related activities) which is usually explained as a transient psychophysiological response to an insufficiently meaningful educational situation" (Finkielsztein 2020, p. 2).

4.2.4 Conceptual boundaries

Boredom can be further delineated by considering conceptually related states that overlap with it in certain respects but remain theoretically distinct.

4.2.4.1 Related concepts

- **Frustration.** Boredom and frustration are both unpleasant states that can signal that the current situation should change, and they may co-occur under excessive task demands. However, Westgate and Wilson (2018) distinguish them by arguing that *boredom concerns the process of engagement*, arising when available cognitive resources do not match task demands, whereas *frustration concerns blocked goal attainment*. Accordingly, task difficulty shows a curvilinear relationship with boredom, which can arise from both under- and overchallenge, but a more linear relationship with frustration, which increases as difficulty increasingly obstructs successful performance (Westgate and Wilson 2018).
- **Disengagement.** Boredom is closely related to disengagement but is not synonymous with it. Disengagement describes reduced behavioural, cognitive, or emotional involvement in an activity, whereas boredom is an aversive affective experience associated with unsuccessful engagement. Consequently, boredom may be an affective component, antecedent, or consequence of disengagement rather than disengagement itself (Eastwood et al. 2012; Skinner et al. 2009; Sharp, Sharp and Young 2020).

4.2.4.2 Contrasting concepts

- **Interest**: Contrary to boredom, *interest* is an "intrinsically valuable […] engagement with an activity, object, or content" (Pekrun, 2019, p. 909). Cognitive interest is heightened by clear, easy-to-understand content (Mazer, 2017:352; Pekrun, 2019), while cognitive overload can also trigger boredom. Both interest and boredom show curvilinear but opposing relationships with task challenge: boredom follows a U-shaped pattern (Westgate and Wilson 2018), while interest shows a bell-shaped pattern, with high values at an optimal level of challenge (Dey and Gottlieb 2019). Boredom, nevertheless, is not only a state without interest: it

is typically characterised by low valence and may be accompanied by emotions such as anger, sadness, or apathy (Goetz et al. 2024). Boredom and interest influence each other over time: boredom may prompt cognitive engagement by encouraging students to seek meaning or novelty (Mugon, Danckert and Eastwood 2019). A recent longitudinal study provided empirical support for this statement, finding that not only does boredom predict lower subsequent interest, but also an increase in interest can reduce boredom in the later stages of students' learning pathways (Boehme et al. 2025).

- **Enjoyment**: Enjoyment can be regarded as an affective contrast to boredom: whereas boredom is an aversive signal that the current activity is failing to engage attention meaningfully and motivates a shift towards alternative activities, enjoyment represents one possible target of such a shift (Westgate and Steidle 2020).
- **Flow**. Csikszentmihalyi describes it as a state 'beyond boredom,' and characterises it in contrast to boredom: Flow emerges when one's skills and task demands are in optimal balance, and the activity is intrinsically rewarding, while boredom arises when challenges are insufficient or mismatched with one's abilities. (Csikszentmihalyi 1975; Weibel and Wissmath 2024). Both conditions are characterised by a distorted perception of time, but for entirely different reasons. Flow theories also highlight that boredom and flow cannot be experienced simultaneously (Csikszentmihalyi, 1990, cited by Wolff et al., 2024). A recent study examining the relationship between the two states (Weibel and Wissmath 2024) nevertheless warns that such a statement would prevent a nuanced understanding of boredom. As mentioned in Section 4.2.1, one important feature of boredom is its evolutionary function: the individual's motivation for uncontrolled exploration, which may even be a precursor to flow.
- **Attention.** One of the universal features of boredom is attentional disengagement. This is due to the original function of boredom: redirecting attention from unrewarding activities to those offering more significant reinforcement and increased cognitive engagement (Pekrun et al., 2014).
- **Engagement**. The present thesis examines boredom in the context of learning and explains boredom as a possible component of disengagement, while maintaining that engagement is a complex and overarching concept with various components,

antecedents and effects (See Chapter 2). Boredom is often defined as "a primary indicator of student disengagement" (Sánchez-Rosas & Esquivel, 2016., p. 2) or a state of disengagement. Academic boredom is negatively correlated with behavioural and cognitive engagement (Goetz et al. 2014), and in the context of learning, emotional interest – feelings of enjoyment, excitement, or fascination – is strongly linked to higher levels of student engagement in both behavioural and cognitive domains. As emotional interest promotes attentional focus and meaningful engagement, its absence may increase the likelihood of disengagement. Boredom is also defined as "involving a motivation to avoid engagement with an activity perceived as aversive or meaningless" (Pekrun et al. 2010, p. 533).

Neither of the above concepts can be considered the direct opposite of boredom; their relationship with boredom contributes to a more comprehensive analysis of the factors needed to prevent boredom. The concepts described in the present section are not only the opposite equivalents of boredom but are also effective *remedies* to prevent or eliminate student boredom: by most intervention strategies, it is easier to increase the level of student enjoyment by expressing teacher enthusiasm rooted in interest, engagement, flow, or intrinsic motivation than striving to overcome boredom because these emotional-cognitive states are "incompatible with boredom" (T. Goetz et al. 2024, p. 239). Strategies for coping and preventing boredom are discussed in detail in Sections 4.2.4 and Chapter 5.

4.2.5 Theoretical accounts of boredom

Without aiming to provide an exhaustive overview, this section highlights several theories that played an important role in the development of the MAC model, which serves as the main theoretical framework for boredom in the present study. Although these theories originate from different disciplines, they engage with and build on one another, creating a degree of continuity between them.

Attentional accounts. Eastwood et al. (2012) explain boredom mainly in terms of difficulties with attentional engagement. In their account, boredom occurs when individuals want to engage in a satisfying activity but are unable to direct or sustain their attention.

Control-Value Theory. Within educational psychology, Control-Value Theory explains academic boredom through students' appraisals of control and value and places boredom within a broader system of achievement emotions (Pekrun 2006; Pekrun and Goetz 2024). Its particular relevance to the present thesis is that it connects boredom to learning, achievement and other academic emotions, as discussed in Section 3.3.1.

Meaning-based accounts. From a social psychological perspective, Van Tilburg and Igou (2012, 2017) foreground perceived meaninglessness as a source of boredom. In this account, boredom signals that the current activity lacks purpose or relevance and may motivate individuals to seek or restore a sense of meaning.

The Meaning and Attention Components model. Westgate and Wilson's (2018) Meaning and Attention Components (MAC) model integrates attentional and meaning-based explanations, proposing that boredom can arise from insufficient meaning or a mismatch between cognitive demands and available attentional resources. Importantly, attentional mismatch may result from both under- and overstimulation. As the MAC model provides the primary boredom framework for the present thesis, it is discussed in greater detail in Section 4.2.6.

Boredom Feedback Model. More recent work has focused on the dynamic processes that follow the experience of boredom. The Boredom Feedback Model conceptualises boredom as feedback indicating that current engagement is unsatisfactory and suggests that this signal can initiate changes that may restore satisfactory engagement.

Functional accounts. While the previous theories mainly focus on the causes of boredom, functional accounts examine its role and consequences (Bench and Lench 2013; Danckert et al. 2018; Elpidorou 2018). They suggest that the unpleasant experience of boredom signals that current engagement is unsatisfactory and motivates the individual to change the activity, goal, or way of engaging with the situation.

The above theories approach boredom from different perspectives and focus on partly different questions. Over time, the focus of boredom research has broadened from particular causes, such as problems with attention or a lack of meaning, towards models that combine several mechanisms and examine the regulatory and functional role of boredom. The theories, therefore, often complement rather than replace one another, while

also differing in several important assumptions; the debates most relevant to the present study are discussed in Section 4.3.

4.2.6 Chosen theoretical framework and working definition of boredom

The MAC model was selected as the primary theoretical framework for interpreting boredom because it explains boredom through learners' situated experience of attentional fit and perceived meaning rather than assuming that particular objective characteristics of a task necessarily produce boredom (Westgate and Wilson 2018; Westgate and Steidle 2020). This appraisal- and meaning-centred account is compatible with the broader epistemic orientation of the present thesis, particularly its attention to learners' interpretations and to variation across individuals and contexts.

In the present thesis, *boredom is conceptualised as a transient, aversive affective state signalling unsuccessful engagement with an ongoing activity, arising when attentional demands and/or perceived meaning are insufficiently aligned with the individual and the situation.*

4.3 Unresolved theoretical questions

This section outlines the most relevant debates to the present study, to which its findings may offer further empirical insight. Some differences between boredom theories may partly reflect their roots in different areas of psychology, which foreground different aspects of boredom, such as attention, affect, motivation, or learning. They may also arise from differences in explanatory focus, research context, and the conceptualisation and operationalisation of key constructs.

4.3.1 Is boredom an emotion?

The classification of boredom has remained theoretically contested. While contemporary research increasingly conceptualises boredom as an emotion (Westgate and Steidle 2020; Tam *et al.* 2021; Pekrun and Goetz 2024), earlier and alternative approaches have described it as a mood, cognitive state, or affective experience. These differences partly reflect broader disagreements about how emotions themselves should be defined and which features are necessary for a psychological state to qualify as an emotion (Westgate

and Steidle 2020). Accordingly, the present thesis does not attempt to determine whether boredom is *in essence* an emotion rather than a cognitive state or mood.

Instead, boredom is treated here as a multicomponent affective state involving interrelated affective, attentional, motivational, physiological, and behavioural processes. Across theoretical traditions, boredom is consistently characterised as an aversive experience associated with unsuccessful engagement, difficulty sustaining attention, a desire to change or leave the current situation, and changes in behaviour and bodily expression (Danckert et al. 2018; Eastwood et al. 2012; Goetz et al. 2014; Westgate and Wilson 2018). Its specific configuration may vary across situations: boredom may involve relatively low or high arousal, different degrees of attentional disengagement, and different behavioural responses rather than a single invariant physiological or expressive pattern.

This multicomponent conceptualisation is particularly appropriate for the present research because it does not assume that boredom can be fully captured by any single indicator. Instead, subjective experience, attentional and motivational processes, and observable affective and behavioural expressions provide complementary information about the state. This perspective, therefore, provides the conceptual basis for the multimodal measurement approach adopted in the thesis, in which self-reports, retrospective accounts, and facial-behavioural indicators are combined to examine how boredom emerges and changes during learning.

4.3.2 Does boredom always involve low arousal?

Theoretical and empirical studies on boredom consistently describe it as a "low-valence" state, i.e., characterised by unpleasant sensations (T. Goetz et al. 2024; Pekrun 2006; Pekrun and Goetz 2024; Russell 1980). Regarding the arousal level during boredom, irrespective of the measurement applied, different studies have reached divergent conclusions:

a) The *low-arousal approach*: The early theoretical papers discussing the nature of emotional states (Russell, 1980) and many studies relying on self-report-based data collection report that boredom is impaired by low arousal attributed to situations that offer inadequate stimulation (Mikulas and Vodanovich 1993; Pattyn et al. 2008; Van Tilburg and Igou 2012; Vogel-Walcutt et al. 2012).

b) The high-arousal approach: A group of studies suggest that boredom is best characterized as an agitated or restless state associated with higher physiological arousal (Berlyne 1960; Jang et al. 2015; Ohsuga, Shimono and Genno 2001).

However, further studies investigating boredom have found that, in some cases, arousal levels during periods of boredom are high, while in others they are low. This phenomenon was explained by different approaches (Merrifield and Danckert 2014; Chin et al. 2017; Westgate and Wilson 2018; Westgate 2020).

c) The *multi-type arousal* approach: The multi-type arousal approach distinguishes five situational boredom profiles based on their relative positions along the valence and arousal dimensions (indifferent, calibrating, searching, reactant, and apathetic boredom) (Goetz et al., 2014). However, the model does not define absolute thresholds for high or low arousal, as these are determined relative to the distribution within a given sample, which limits its suitability for measurement and comparison across studies. The model is therefore primarily descriptive, which may also help explain why it has not become a widely used framework for operationalising boredom.

Approaches a), b) and c) suggest that the level of arousal in a given boredom situation may be influenced by a combination of circumstances and personality traits (Finkielsztein 2020; Wolff, Radtke and Martarelli 2024)

d) The *arousal discrepancy* approach: Another group of studies suggests that the state of boredom can include high and low arousal states, even at the same time: it can occur that the individual experiences a subjective state of frustration and dissatisfaction while their objective neurological state shows low arousal (e.g., Danckert et al. 2018; Eastwood et al. 2012; Fahlman et al. 2013; Merrifield and Danckert 2014). Studies applying physiological measures (heart rate and skin conductance) found that boredom is characterized by increased arousal (Merrifield and Danckert 2014).
e) The *arousal-independent approach*: The above theories have been challenged by studies arguing that boredom should be considered a ubiquitous state which can be described by motivational and cognitive factors such as *perceived meaninglessness* and *level of attention*, instead of valence and arousal levels (Elpidorou 2021; Van

Tilburg and Igou 2012; Westgate and Wilson 2018; Wolff, Radtke and Martarelli 2024). Others pointed out that the concept of arousal is not clearly defined: if boredom is multidimensional, different measures can assess only some of its dimensions (Watson and Tellegen 1985).

The unresolved question is therefore not whether boredom has a single characteristic arousal level, but whether variation in arousal reflects different boredom states, distinct underlying mechanisms, individual differences, or differences between subjective and physiological/behavioural indicators.

4.3.3 What makes an activity boring: difficulty, lack of meaning, or both?

One of the central unresolved questions concerns whether boredom results mainly from insufficient stimulation or from excessive cognitive demand. Early flow-based accounts associated boredom primarily with underchallenge, in which task demands fall below an individual's capacities, engagement decreases, and boredom becomes more likely. Later research, however, suggests that this relationship is not simply linear. Westgate and Wilson (2018) found that boredom was higher both when cognitive demands were too low and when they were too high, and lowest when demands were better matched to available mental resources. In contrast, frustration increased more continuously with task difficulty, suggesting that boredom under conditions of overchallenge cannot be reduced simply to frustration.

The MAC model explains this pattern through attentional fit rather than task difficulty alone. As shown in

Table 3, the attentional component distinguishes four conditions: understimulation, low-level engagement, high-level engagement, and overstimulation. Successful engagement can therefore occur under both low and high demand, provided that available resources match those demands. Boredom emerges when this fit breaks down, whether because demands are too low or because they exceed available resources.

Table 3 - Profiles of Boredom and Their Predicted Outcome as a Function of Meaning and Attention

Attention Component	Meaning Component: Low Meaning: Task is INCONGRUENT with Valued Goals	Meaning Component: High Meaning: Task is CONGRUENT with Valued Goals
Understimulation: Demand < Resources	A — ***Meaningless + Attentional Boredom*** *Seek Interesting Activity*	E — ***Attentional Boredom*** *Increase Demand*
Low-Level Engagement Low Demand + Low Resources	B — ***Meaningless Boredom*** *Seek Enjoyable Activity*	F — ***Enjoyment (Low Boredom)***
High-Level Engagement High Demand + High Resources	C — ***Meaningless Boredom*** *Seek Interesting Activity*	G — ***Interest (Low Boredom)***
Overstimulation: Demand > Resources	D — ***Meaningless + Attentional Boredom*** *Seek Enjoyable Activity*	H — ***Attentional Boredom*** *Increase Resources*

Note. *Cells are lettered for ease of reference.*

Source: (Westgate and Wilson 2018, p. 17)

A related unresolved question concerns emotional specificity: when high task value is combined with excessive challenge, does attentional mismatch itself produce boredom, as the MAC model allows, or does high value shift the emotional response towards anxiety or frustration, as predicted by CVT?

The MAC model provides relatively direct evidence that excessive cognitive demand can increase boredom rather than merely producing frustration. Westgate and Wilson found a curvilinear relationship between cognitive demand and boredom: boredom was higher when demands were either too low or too high, and lowest at moderate levels of demand. In contrast, frustration increased much more continuously and steeply as demand increased, suggesting that boredom under overchallenge is not simply frustration labelled differently. This supports the MAC claim that attentional mismatch can generate boredom through both understimulation and overstimulation. However, the more specific prediction

that boredom also increases under high demand, combined with high meaning/value, has not been tested as directly. This is important for the comparison with CVT, which predicts that a high value should protect against boredom even under overchallenge. Thus, the strongest unresolved issue is not whether overchallenge can produce boredom, for which the MAC has direct evidence, but whether high value prevents or reduces this overchallenge-related boredom.

4.3.4 Two pathways for alleviating boredom: interest and enjoyment

Boredom creates a strong motivation to change the current state. Westgate and Wilson (2018) argue that one way to alleviate boredom is to switch to another activity, but what people switch to matters as well. They distinguish two possible pathways: people may seek experiences providing either *interest* or *enjoyment*.

- *Interest* is typically supported by novelty, complexity, and uncertainty and requires more cognitive effort,
- whereas *enjoyment* is more closely associated with familiar, simple, and positively rewarding experiences and requires less processing.

The two pathways may therefore move in opposite directions: factors that increase interest may reduce enjoyment, and vice versa. As illustrated in Figure 6, repeated reliance on *enjoyable* rather than *interesting* alternatives may relieve short-term boredom but offer fewer opportunities to develop new interests, thereby increasing susceptibility to boredom over time.

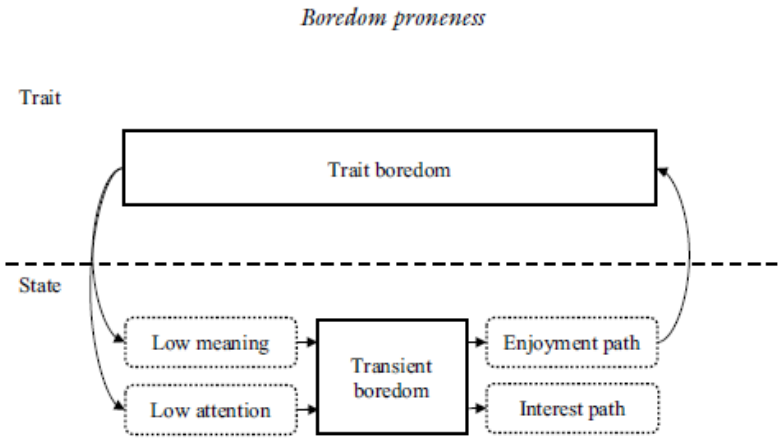


Figure 6: Trait boredom and the meaning and attention model.
Source: Tilburg *et al.* 2024

4.4 Strategies to reduce boredom

In general, as in the case of other emotions, it is easier to prevent boredom than to reduce it once it has emerged. The second principle, which applies to most intervention methods, is that to overcome existing boredom, it is more effective to *foster certain positive and activating emotions* than to suppress distracting thoughts (Frenzel et al. 2018; Herrando and Constantinides 2021; Pekrun and Goetz 2024).

Using the logic and the terminology of the Control-Value Theory (see Section 3.3.1), three types of situations can trigger boredom:

(a) *low control* and *low values* (e.g., when the student does not feel the task meaningful and has a low level of perceived autonomy);
(b) *low control* (e.g., perceived task overchallenge), and
(c) *high control* and *low values* (e.g., perceived underchallenge, in the case of a meaningless task) (Tze, Parker and Daniels 2024).

One of the key elements of the learner-centred approach in constructivist pedagogy is learners' responsibility for self-regulated learning. While the present thesis's length and scope allow only a limited exploration of student coping strategies, this in no way implies that these strategies are considered less important than teacher-led interventions.

4.4.1 Coping with boredom as a student

Although boredom has a stronger effect on students' academic results than anxiety (Daniels et al., 2009; Tze et al., 2016, as mentioned by Tze et al., 2024), coping strategies for students' boredom have been researched far less than its causes and precursors.

A recent framework, compiled by Tze et al. (2024), has been developed based on (a) the review of the research on the subject, (b) applying the framework of Control-Values Theory and (c) integrating the main elements of the psychological practices that have been developed for cognitive and emotional interventions applied for coping with stress, negative emotions, and thoughts. As boredom is associated with perceived control and value, the five mechanisms supporting adaptive boredom regulation proposed by the above

study are not necessarily focused on boredom itself but on the affective, cognitive, behavioural and situational components or triggers related to it:

- Step 1: *Psychoeducation* (a cognitive-behavioural approach): the first and most important step is to make students aware that boredom during learning is normal, but also a serious risk, and can be successfully managed with appropriate strategies.
- Step 2: *Cognitive restructuring* (a cognitive-behavioural approach): identifying and changing unhelpful thought patterns that maintain boredom and modify problematic behaviours related to it. According to the CVT, the antecedents of boredom are associated with perceived control and value; this restructuring can increase perceived control by driving learners' attention from the significance of their "abilities" to the importance of their "efforts". This strategy is similar to attributional retraining interventions, based on the principle of attribution theory, which states that how we frame important outcomes and what causes we attribute to them strongly impact how we feel, think, and behave (Parker et al. 2018).
- Step 3: *Situation modification* (emotional-regulation strategy); i.e., within the possibilities available in the given learning context, students can change their behaviours or some of the circumstances of the problematic situation in a way that reduces negative emotions related to the completion of the challenging task while following an *approach* strategy towards it (e. g., by sitting in the front of the lecture hall, contacting the teacher with their questions after class, or finding value in the process, etc.).
- Step 4: To prevent and tackle mind wandering, *attention deployment* (an emotional-regulation strategy) can include (a) information sharing on the effects of inattention and (b) help students with identifying a set of personalised adaptive methods that can be used to direct their attention back to the learning content (Harley et al. 2019).
- Step 5: *Relapse prevention* (a cognitive-behavioural approach), i.e., students need to be prepared to identify their boredom triggers, and by planning their learning activities based on this awareness, they can reduce the risk of slipping back into a state of boredom (Tze, Parker and Daniels 2024).

The Boredom Feedback Model also highlights the importance of adaptive coping. Strategies that repeatedly provide quick relief from boredom, such as turning to a

smartphone, may become reinforced and increasingly habitual; therefore, coping should ideally help learners restore engagement in ways that support their longer-term goals (Tam et al. 2021).

4.4.2 Instructional interventions to prevent and reduce boredom

Instructors have a wide range of powerful ways to prevent and reduce boredom, which is outlined below, along with the main elements of Control-Value Theory (CVT).

While the reciprocal relationship between academic emotions and engagement can foster negative spirals of disengagement, (Pekrun and Linnenbrink-Garcia 2022) highlight an encouraging perspective: This circularity also offers powerful opportunities for positive change, suggesting that well-designed interventions can successfully reinforce upward spirals of engagement and emotional wellbeing (referring to Artino Jr. and Stephens 2009; Ben-Eliyahu et al. 2018; Harley et al. 2019), as follows: 'since emotions, their antecedents, and their effects can be reciprocally linked over time, emotions can be regulated and changed by addressing *any of the elements* involved in these cyclic feedback processes' (Pekrun & Linnenbrink-Garcia, 2022, p. 127, *emphasis added*). Such interventions can address

- emotions;
- control and value appraisals;
- competences related to agency (e.g., learning skills);
- or tasks and learning environments.

These are in line with the basic needs related to intrinsic motivation, namely, *relatedness*, *competence* and *autonomy* as highlighted by the Self-Determination Theory (Ryan and Deci 2000). Pekrun and Goetz summarised the interrelationship between the two theories as follows:

> 'If environments ease actions that fulfil needs for competence, autonomy, and relatedness […], these actions become valuable, thus preventing boredom. From a theory perspective, this is a point where CVT and self-determination theory meet: *Self-determination theory explains the generation of value; CVT explains how value generates emotions.*' (Pekrun & Goetz, 2024, pp. 80–81, emphasis added)

4.4.2.1 Control-related interventions

If the course design and delivery (a) allow learners a certain degree of autonomy in the choice of topics, types of tasks, or pace of certain learning activities, (b) develop their competencies, and (c) ensure appropriate alignment between students' abilities and task challenges, students' perceived control increases, which in turn helps prevent or reduce boredom. Teachers can develop their diagnostic skills to detect increases in student boredom and adjust their activities accordingly. The measures serving the clarity of teachers' explanations (discussed in Chapter 5) prevent boredom by reducing the risk of perceived overchallenge.

4.4.2.2 Value-related interventions

Instructors can increase the perceived *value of the learning content*, by "value messages", explicitly referring to the usefulness and the relevance of the given content or activity in terms of their everyday life or in the course of the assessment. The implicit expression of the teacher's enthusiasm and behaviour is nevertheless a much more powerful way to increase the value of the content discussed in the course: as a result of the *emotional contagion*, the teacher's passion and enthusiasm induces an immediate sense of enjoyment and excitement in students (Frenzel et al. 2018; Herrando and Constantinides 2021; Pekrun, Marsh, Elliot, et al. 2023).

Instructors can increase the perceived value of learning by enhancing the meaningfulness of *interactions* within the learning environment. Through their immediate and encouraging communication, they can create a social presence – even in the context of online learning – that fosters intrinsic motivation by contributing to the fulfilment of students' basic psychological needs (Garrison, Anderson and Archer 1999; Gorham 1988; Pekrun and Goetz 2024, p. 202; Ryan and Deci 2000). The effects of interventions that enhance the teacher-student relationship will be discussed further in Chapter 6. The antecedents, consequences, and prevention strategies identified by the review are presented in Figure 7.

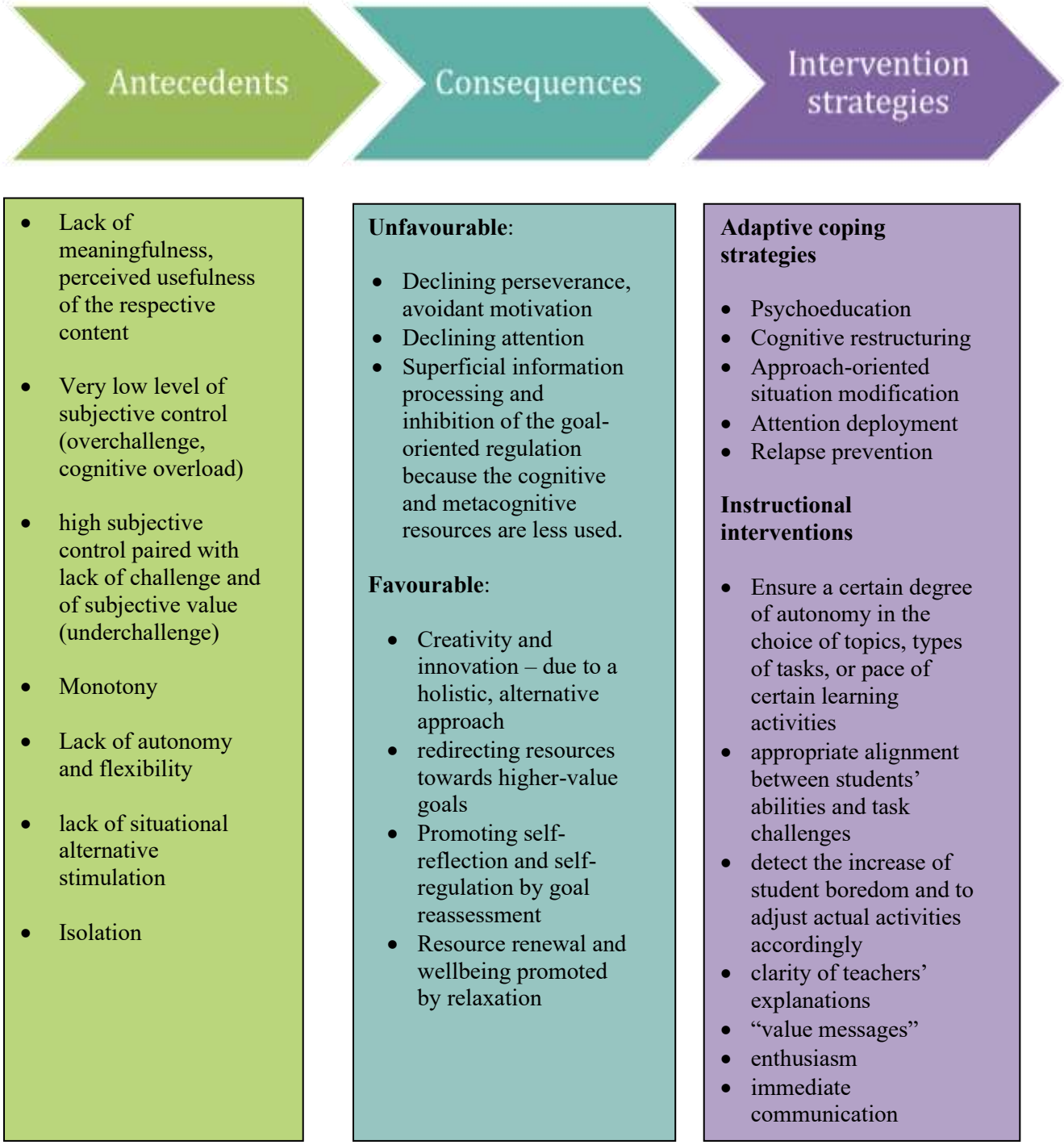


Figure 7: Antecedents and consequences of boredom.

Author's own compilation, based on Goetz, Hall and Krannich 2019a; Pekrun and Goetz 2024; Sass, Vinczéné Fekete and Bodnár 2023; Van Tilburg and Igou 2012)

5 THE POTENTIAL OF INSTRUCTIONAL COMMUNICATION

Figure 7 lists the main instructional interventions that can increase student engagement and reduce deactivating emotions such as boredom. This chapter narrows its focus from all of these to *instructional communication.* Drawing on instructional communication research and related theoretical perspectives, this chapter examines how rhetorical and relational aspects of instructional communication may support students' cognitive and affective engagement. It focuses particularly on teacher clarity and immediacy as complementary communication qualities and establishes the broader conceptual framework that will later be applied to instructional-video features in Chapter 6.

Scientific examination of *Instructional communication* within the discipline of communication science has a relatively short history, as the first institutionalised initiative to the systematic examination of instructional communication was established only in 1972 when the Instructional Communication Division was formed within the International Communication Association (Farris, Houser and Hosek 2017), but the roots of exploring factors of effective instructional communication date back much further.

The first part of the chapter reviews selected theoretical distinctions that have contributed to contemporary understandings of the cognitive, affective, rhetorical, and relational dimensions of teaching and learning. These include the cognitive and affective domains of educational objectives, rhetorical and relational perspectives in instructional communication, teacher clarity and immediacy, and learner-side appraisal processes described by Control-Value Theory. The second part considers how these concepts apply to mediated and online instructional environments and prepares the conceptual groundwork for the asynchronous instructional video examined in Chapter 6.

5.1 Rhetorical and relational perspectives in instructional communication

As discussed in Chapter 2, early research on student engagement frequently concentrated on observable behavioural indicators, such as participation and time-on-task. Later conceptualisations increasingly recognised cognitive and affective dimensions of engagement alongside behavioural engagement. Instructional communication research approaches these processes from the perspective of the verbal and nonverbal messages

exchanged between teachers and learners and the meanings and responses associated with them.

Instructional communication has commonly distinguished between rhetorical and relational perspectives (Mottet and Beebe 2006; Houser and Hosek 2022). The rhetorical perspective foregrounds instructional messages and the instructor's attempts to influence learners' understanding, attitudes, beliefs, or behaviour. Relational approaches foreground the interpersonal dimension of teaching and the role of communication in establishing and maintaining teacher–student relationships. These perspectives are complementary rather than mutually exclusive.

Mottet and Beebe (2006) further distinguish communication-as-action, communication-as-interaction, and communication-as-transaction. Communication-as-action represents a predominantly one-way, message-centred process. Communication-as-interaction incorporates feedback, while communication-as-transaction emphasises reciprocal meaning-making. This distinction becomes particularly relevant to asynchronous instructional video, which is structurally closest to communication-as-action because the learner cannot provide feedback that changes the message during viewing. Nevertheless, a one-way instructional message can still contain relational cues and communicate instructor presence, approachability, and audience orientation.

A distinction between cognitive and affective aspects of learning is also present in Bloom's taxonomy of educational objectives and the subsequent taxonomy of the affective domain (Bloom 1956; Krathwohl, Bloom and Masia 1964). Cognitive objectives concern intellectual processes and knowledge, whereas affective objectives concern interests, attitudes, values, and related orientations. In the present framework, this distinction is included as an example of the long-standing recognition that learning involves both cognitive and affective dimensions. It is not treated as equivalent to the rhetorical–relational distinction in instructional communication.

Bolkan and Griffin (2018) distinguish instructional interventions by their contributions to "catching" and "holding" students' situational interest. Catching interest involves attracting and sustaining attention through features such as immediacy, humour, and intellectual stimulation, whereas holding interest is more strongly associated with relevance and meaningfulness, as well as the development of autonomous motivation. The

distinction is relevant to the present study because it demonstrates that instructional communication may support engagement through different attentional and motivational pathways. However, catch and hold do not map directly onto rhetorical and relational communication and are therefore not treated as equivalent dimensions in the framework below.

The Cognitive Theory of Multimedia Learning (CTML; Mayer 2021) provides a predominantly cognitive account of how learners process words and pictures. Multimedia instructional principles are organised around three major instructional goals: reducing extraneous processing, managing essential processing, and fostering generative processing. Generative processing should not be interpreted as an affective or relational category; it refers to the cognitive activity involved in making sense of instructional material. Later multimedia-learning research has increasingly incorporated motivational, social, and affective processes, creating areas of closer contact with instructional communication. These developments are considered in detail in Chapter 6.

Table 4: Selected concepts reflecting rhetorical and relational emphases in instructional communication

Source / theoretical or conceptual framework	Analytical focus	Predominantly rhetorical/cognitive orientation	Predominantly relational/affective orientation
Bloom (1956); Krathwohl, Bloom and Masia (1964)	Learning objectives	**Cognitive** learning objectives	**Affective** learning objectives
Rhetorical/Relational Goal Theory – Mottet, Frymier and Beebe (2006)	Instructor goals	Rhetorical goals	Relational goals
Rhetorical/Relational Goal Theory – Mottet, Frymier and Beebe (2006)	Student needs	Academic needs	Relational needs
Teacher clarity and immediacy research – Andersen (1978); Mehrabian (1981); Mottet and Beebe (2006); Mazer (2022)	Instructional communication behaviours	Teacher clarity: organisation and comprehensibility of instructional messages	Teacher immediacy: communication reducing psychological distance
Cognitive Theory of Multimedia Learning – Mayer (2021)	Multimedia learning and instructional design	Reducing extraneous processing; managing essential processing	No direct relational equivalent; motivation and affect enter the framework through later extensions of multimedia-learning research
Control-Value Theory – Pekrun (2006)	Learner appraisal	Perceived control	Perceived value

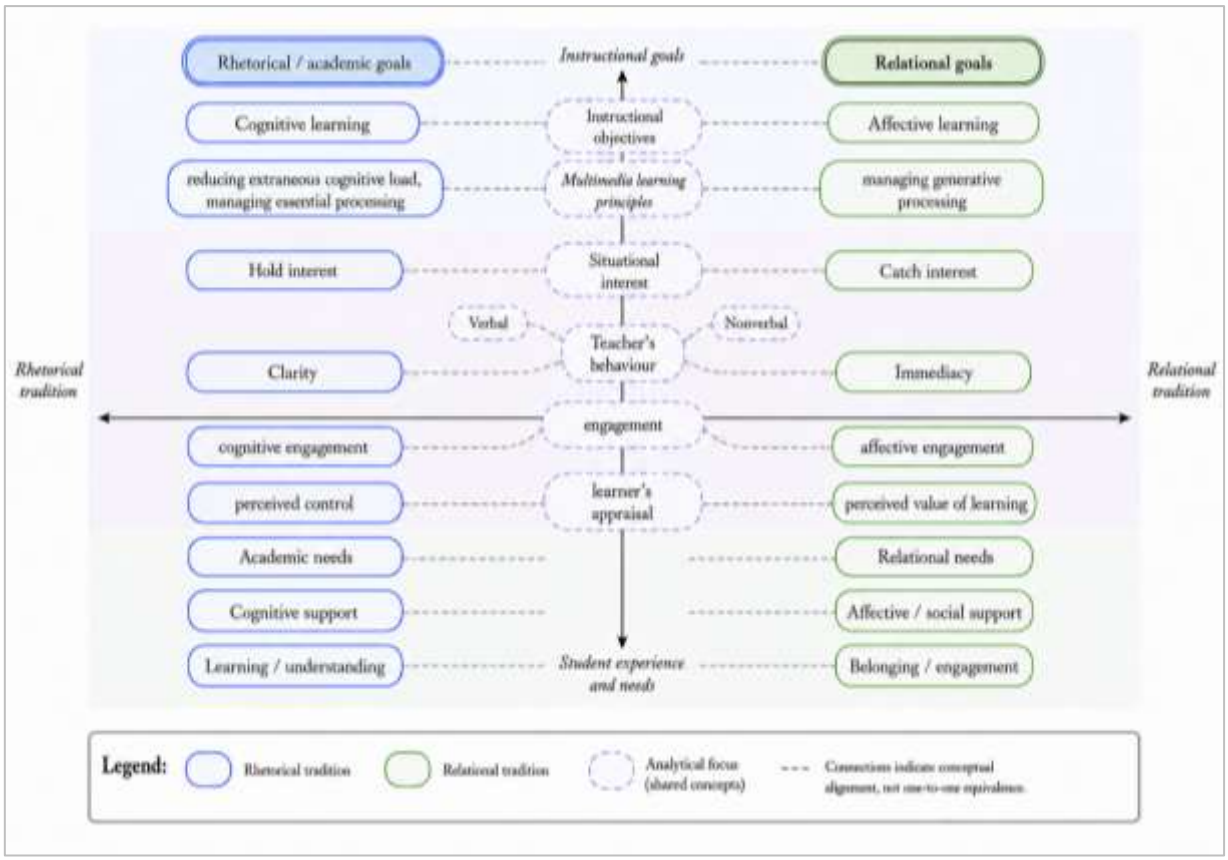
Note. Author's own conceptual synthesis based on the sources cited in each row. The positioning indicates predominant conceptual affinity rather than equivalence, synonymy, or mutually exclusive categories. The

frameworks operate at different analytical levels, including learning objectives, instructor goals, student needs, communication behaviours, cognitive processing, and learner appraisals.

The concepts summarised in Table 4 operate at different analytical levels and should therefore not be interpreted as interchangeable labels. Their relationship is represented more clearly in Figure 7. The vertical dimension of the model distinguishes the instructional side of the teaching–learning process from the learner-experience side. Instructor goals, communication behaviours, and instructional design are positioned towards the instructional end, whereas student needs and appraisals are positioned towards the experiential end. The horizontal dimension represents a predominant orientation from rhetorical or broadly cognitive functions towards relational or broadly affective functions.

The model is anchored in Rhetorical/Relational Goal Theory (Mottet, Frymier and Beebe 2006), which distinguishes instructors' rhetorical and relational goals and students' academic and relational needs. The other concepts are positioned around these distinctions according to their primary conceptual affinity. Their placement does not imply that cognitive and affective, rhetorical and relational, or clarity and immediacy are interchangeable categories.

Figure 8: A four-dimensional model of rhetorical and relational traditions in instructional communication



Note. Author's own conceptual synthesis based on Mayer (2021), Mazer (2022), Mottet, Frymier and Beebe (2006), and Pekrun (2006).

The categories represented in Figure 7 indicate predominant orientations rather than discrete divisions. Cognitive and affective processes are closely interconnected, as are rhetorical and relational functions in instructional communication. A communication behaviour that primarily supports understanding may also communicate attentiveness or care, while a behaviour primarily associated with relational closeness may simultaneously attract attention or support comprehension. In the present thesis, rhetorical and relational therefore refer to broader communicative orientations, whereas clarity and immediacy are used as more specific, non-exclusive lenses for interpreting instructional communication behaviours.

The MAC model likewise cuts across the rhetorical–relational organisation of the present framework. Attentional fit cannot be equated with rhetorical communication, and perceived meaning cannot be equated with relational communication. Both rhetorical and relational features may influence attention and meaning. For this reason, the MAC model

is used as a complementary explanation of boredom and is not included among the dimensions in Figure 7.

5.2 Teacher immediacy and clarity

Within the broader rhetorical–relational framework, the present thesis uses teacher clarity and teacher immediacy as more specific analytical concepts. Clarity refers primarily to communication behaviours that support the organisation, comprehensibility, and fidelity of instructional messages, whereas immediacy refers primarily to behaviours that reduce perceived psychological distance and communicate interpersonal approachability. The distinction is analytical rather than absolute: both may contribute to improved attention, emotional responses, motivation, and engagement.

5.2.1 Immediacy

Both verbal and nonverbal communication can perform rhetorical and relational functions. Nevertheless, instructional immediacy research has given particular attention to nonverbal behaviours such as eye contact, smiling, bodily orientation, gesture, and vocal expression because these behaviours can communicate psychological closeness and approachability with relatively little explicit verbal content (Andersen 1978; Mehrabian 1981; Mottet and Beebe 2006).

One of the most extensively studied variables in instructional communication is immediacy. Mehrabian (1981) used the term to describe communication behaviours associated with psychological approach and preference. In instructional settings, immediacy encompasses verbal and nonverbal behaviours that may reduce the perceived distance between the instructor and the learner.

Emotional Response Theory (ERT; Mottet, Frymier and Beebe 2006) proposes that instructional communication behaviours and conditions can elicit emotional responses that are related to subsequent approach or avoidance tendencies towards the instructor, the learning situation, and learning itself. An important implication of the theory is that instructional communication cannot be evaluated solely on the basis of the instructor's behaviour; learners' emotional responses to that behaviour also matter.

Contemporary evidence suggests that teacher immediacy behaviours are positively associated with student participation, cognitive engagement, and learning outcomes in university settings (Amini and Weisi 2023; Liu et al. 2021), and this association is particularly relevant in the case of complex subjects where motivation is essential for persistence (Titsworth et al. 2013).

5.2.2 Clarity

Clarity is defined as a cluster of teacher behaviours that contributes to the fidelity and comprehensibility of instructional messages (Mazer 2022). Clarity behaviours include signalling important information, organising explanations, providing examples and summaries, and making relationships among ideas explicit. Although clarity is predominantly rhetorical and message-oriented, its consequences are not exclusively cognitive. Research has associated teacher clarity with students' cognitive interest and emotional responses to instruction.

Immediacy aligns more closely with the relational tradition of teacher communication, while 'clarity' reflects the rhetorical tradition. Through the lens of the rhetorical tradition of communication (Craig 1999), teaching can be defined as an instructional action of sending "messages to influence or persuade students with the goal of either changing or reinforcing their attitudes, beliefs, values, or behaviors" (Farris, Houser and Hosek 2017, p. 7). The Rhetorical/Relational Goal Theory (Mottet, Frymier and Beebe 2006) holds that teachers usually have both *rhetorical* and *relational* goals, and that students, similarly, have *academic* and *relational* needs that can vary across personalities and situations. The teaching-learning process is effective when instructors "set appropriate rhetorical and relational goals, and […] are able to use appropriate communication strategies to accomplish their goals" (Mottet, Frymier and Beebe 2006, p. 268) and "when students' academic needs and relational needs are met" (Mottet, Frymier and Beebe 2006, p. 269). As Mottet (2015) summarised it, "cognitive and affective learning are so closely connected and interdependent that separating them is an artificial bifurcation that is no longer theoretically valid or empirically supported" (p. 508, cited by Waldbuesser and Titsworth 2022, p. 172.). The *cognitive appraisal theory*, emphasising the interplay between cognitive evaluations and emotional responses (Lazarus 1991), highlights that the way students perceive and interpret educational experiences significantly influences their emotional reactions and coping mechanisms, thereby impacting learning outcomes

(Pekrun, 2006). Immediacy can therefore be located closer to the relational orientation of instructional communication and clarity closer to its rhetorical orientation, but neither belongs exclusively to one domain. RRGT explicitly assumes that instructors simultaneously pursue rhetorical and relational goals and that students differ in the relative importance of their academic and relational needs.

Though several studies have explored effective communication tools to enhance teacher immediacy and clarity and to catch and hold student interest in general, there has been little research on communication tools applicable to instructional videos for the same purposes (Tan Ti-Shen and Sen 2019; Woolfitt 2015).

5.2.3 Control and value as learner-side appraisals

Control-Value Theory provides a learner-side perspective that complements the instructor-oriented communication concepts discussed above. CVT does not divide instructional communication into rhetorical and relational categories. Instead, it proposes that achievement emotions depend substantially on learners' appraisals of control and value (Pekrun 2006). Communication and instructional design may contribute to these appraisals, but control and value remain learner interpretations rather than properties of individual teacher behaviours.

Clear organisation, guidance, appropriate difficulty, and comprehensible explanations may contribute to learners' sense of control. Perceived value may be supported through relevance, usefulness, intrinsic interest, personal meaning, and social dimensions of the learning environment. The relationship is nevertheless not one-to-one: the same instructional feature may be appraised differently by different students.

5.3 Instructional communication fostering cognitive and affective engagement in online learning

The relationship between rhetorical and relational communication becomes particularly visible in online learning, where many cues available in face-to-face interaction are reduced, altered, or mediated by technology. This is particularly important in asynchronous environments, in which instructors cannot immediately adapt their explanations to learners' nonverbal feedback and students have fewer opportunities for

spontaneous interpersonal contact which increases the risk of learning failures, and ultimately, student dropout.

Research on online immediacy has identified a range of communicative and design cues through which instructor presence and approachability may be conveyed in mediated environments. Dixson et al. (2017), for example, examined audiovisual content, instructor-related imagery, typography and colour, use of the medium, and chronemic characteristics such as response timing and frequency. Related e-immediacy research has examined personalised communication, instructor self-disclosure, conversational language, timely responses, and other behaviours communicating approachability (Dixson et al. 2017; Ghamdi, Samarji and Watt 2016; Sung and Mayer 2012).

The theory of the *community of inquiry* (CoI, Garrison 2018) has been developed from a study examining the essential conditions for effective educational transactions in online learning, especially asynchronous learning, where social cues such as body language, eye contact, and nonverbal feedback cannot support teacher-student communication. The CoI framework defined three crucial elements of effective computer-mediated communication in education (Garrison, Anderson and Archer 1999a): *teaching presence*, *social presence*, and *cognitive presence*. Of the three, social presence plays a specific role, as it allows participants to experience one another as "real people" in the online environment; together with teaching presence, it constitutes the CoI framework's aspect most closely related to immediacy and interacts with teaching and cognitive presence in supporting the educational experience.

With the unexpected increase in online learning due to the pandemic, the CoI framework, and within it, social presence gained increased attention and research focus. Studies conducted during this period revealed additional strategies to enhance students' social presence, such as turning on their cameras, organising small-group activities via breakout rooms, and providing opportunities for instant chat during lessons. Teachers' behaviours that students appreciated and mentioned as increasing their engagement included humour, commitment, and personal caring (Jia et al. 2020).

Asynchronous instructional videos represent an especially constrained form of mediated instructional communication. They do not allow genuine interaction or immediate adaptation to learner feedback within the video itself. Nevertheless, instructors can

incorporate verbal, visual, and behavioural cues that communicate audience orientation, presence, approachability, and relational intent. At the same time, message structure and audiovisual design must guide learners' attention and comprehension without the possibility of immediate clarification. Chapter 6 therefore examines how multimedia-learning principles and instructional-communication concepts can be brought into dialogue at the level of concrete, observable video features.

6 INSTRUCTIONAL VIDEOS

Instructional videos constitute a distinctive form of mediated instructional communication. In asynchronous video, the instructor cannot immediately recognise declining attention, misunderstanding, or emotional disengagement and cannot modify the explanation in response. Features that support attention, comprehension, motivation, and instructor presence must therefore be incorporated into the instructional artefact before viewing.

This creates an important point of connection between multimedia-learning research and instructional communication. Multimedia learning provides theoretically grounded and comparatively well-defined categories for describing the cognitive design of verbal and visual instructional material, whereas instructional communication provides complementary concepts for understanding the rhetorical and relational functions of teacher messages and behaviours. The two traditions are not equivalent, but selected concepts show sufficient functional overlap to support a feature-level mapping for asynchronous instructional videos.

6.1 Factors influencing the effectiveness of instructional videos

Instructional videos have become an established component of higher education, particularly in blended, distance, and flexible learning environments. Their effectiveness depends not only on the accuracy of the instructional content but also on how information is organised, presented, paced, and communicated to learners. Despite the return to predominantly face-to-face teaching, educational videos are expected to be increasingly used in higher education, especially in flexible learning formats such as blended and distance education. Understanding and measuring student engagement in this environment is therefore important for both instructional design and research.

A particular challenge is that “effectiveness” can refer to different outcomes. Multimedia learning research has focused strongly on retention, transfer, comprehension, and cognitive processing, whereas instructional communication research also examines attention, emotional interest, immediacy, social presence, and engagement. The present thesis, therefore, approaches instructional-video design from both cognitive and communicative perspectives.

6.2 The Cognitive Load Theory and the Cognitive Theory of Multimedia Learning

Cognitive Load Theory (CLT) explains learning in relation to the limited capacity of working memory and the cognitive demands imposed by instructional material (Sweller et al. 2011; Sweller et al. 2019). Contemporary accounts distinguish intrinsic cognitive load, arising from the complexity of the material relative to the learner's prior knowledge, from extraneous cognitive load, generated by unnecessary aspects of instructional presentation. Effective instructional design therefore seeks to minimise unnecessary processing demands and organise learning material so that available cognitive resources can be directed towards constructing knowledge.

Closely related to Cognitive Load Theory and other cognitive theories of learning, the Cognitive Theory of Multimedia Learning (CTML) explains how learners construct knowledge from words and pictures (Mayer 2021). CTML assumes that learners process verbal and visual-pictorial information through partly separate channels, that each channel has limited processing capacity, and that meaningful learning requires active cognitive processing. Learners select relevant information, organise it into coherent verbal and pictorial representations, and integrate these representations with one another and with prior knowledge (Mayer 2021).

Mayer's commonly cited set of 15 multimedia-learning principles was developed from research across several forms of multimedia instruction and was not specific to instructional video.

Fyfield et al. (2022) reviewed 113 peer-reviewed empirical studies published between 1992 and 2021 that examined educational-video design and identified 28 principles investigated specifically in this medium. The reviewed studies primarily assessed outcomes such as retention and transfer. The review showed that empirical support was uneven across principles: some were examined relatively frequently and produced comparatively consistent positive findings, whereas others showed mixed or context-dependent results or had received little empirical investigation. Table 5 summarises the level and consistency of empirical support reported for instructional-video design principles in the review by Fyfield et al. (2022).

Table 5: Empirical support for instructional-video design principles identified by Fyfield et al. (2022)

A: Principles with relatively consistent empirical support	B: Principles with mixed or context-dependent findings	C: Principles with limited empirical investigation in instructional-video research	
Coherence: Including just the essential information Integrating **learning activities**: Make students complete learning activities in addition to watching the video **Embodiment:** include human movement or gestures **Learner control:** Students should be given control over pause, playback, and play-forward. Video length reduction **Segmenting**: Longer videos should be broken into meaningful chunks.	**Modality** (refuted by most studies): Use spoken narration rather than written text **Redundancy** (refuted by most studies): Written text should not be added when narration is present. **Presenter's face**: Avoid including the presenter's face when alternative visuals are displayed. **Signalling**: Important information should be highlighted to learners. **Personalisation**: Narrations should use first/second person conversational speech **Transience**: Video loses advantages over static media when too much information is presented too quickly. **Worked examples**: Include completed guidance or examples when solving problems or learning skills.	Principles considered effective but examined by few studies **Reviews** (3): Videos should end with a summary of the content. **Pre-training** (2): Using words and pictures rather than words alone. **Multimedia** (2): Using words and pictures rather than words alone. **Misconceptions** (2): Conceptual videos dispel common misconceptions at the start. **Spatial contiguity** (3) Related elements presented in close physical proximity on the screen (also called split attention). **Perspective** (1): Videos shot from the learner's perspective are more effective than the third-person perspective. Dialogue (2) Interactivity (4) Principles with mixed or context-dependent findings **Guided discovery** (1): hints and feedback as the learner solves problems. **Emotional design** (4): Warm, high-saturation colours and anthropomorphisms used in videos. **Sound effects** (1): to be avoided due to their distracting effect. **Background music** (3): to be avoided due to its distracting effect. **Speech rate** (2) (fast superior) Audio quality (2)	Principles examined by no empirical research for instructional videos **Voice principle:** Human superior to machine voice. **Drawing**: Asking students to create drawings while reading the text **Temporal contiguity:** Related elements (e.g., narration and visuals) presented at the same time.

Note. Author's own compilation based on Fyfield et al. (2022), adapted from Sass et al. (2023). The categories summarise the quantity and direction of evidence reported in the reviewed instructional-video literature and should not be interpreted as establishing universal effectiveness or ineffectiveness.

6.3 Connecting multimedia learning and instructional communication

Multimedia-learning research and instructional communication address overlapping aspects of instruction but differ in their theoretical origins, terminology, and principal units of analysis. Multimedia-learning principles provide comparatively specific design categories that can be identified in instructional materials, whereas instructional communication offers a broader account of how teacher messages and behaviours function rhetorically and relationally.

The two approaches are therefore not treated as equivalent. Nevertheless, selected constructs show functional overlap. Features such as coherence, signalling, segmenting, and coordinated verbal–visual presentation perform functions closely related to teacher clarity because they support message organisation, direct attention, and facilitate comprehension. Personalisation and instructor embodiment show partial connections with immediacy because they can activate social responses, communicate instructor presence, and reduce perceived psychological distance.

Important differences remain. Instructional communication examines constructs such as humour, smiling, rapport, confirmation, instructor credibility, power, and affinity-seeking that do not correspond directly to individual multimedia-learning principles. Conversely, multimedia-learning research includes principles such as learner control, self-explanation, immersion, and generative activities that have no straightforward equivalents among instructional communication variables.

Table 6: Illustrative areas of overlap and divergence between instructional communication variables and multimedia-learning principles

Conceptual proximity	**Instructional communication behaviours, variables and constructs**	**Multimedia-learning principles**
Close overlap	Clarity / message comprehensibility	Coherence, signalling, segmenting, multimedia
Partial overlap	Immediacy / affinity-seeking	Personalisation, voice, image, embodiment
Additional partial overlap	Confirmation, rapport, disclosure, paralinguistic cues	Feedback, dialogue, collaboration, personalisation, speech rate
Features distinguished primarily in instructional communication	Instructor credibility; power/influence; compliance-gaining; humour; nonverbal immediacy cues such as smiling	-
Features distinguished primarily in multimedia learning	-	Learner control; self-explanation; immersion; generative activities

Note. Author's own compilation.

The comparison does not imply equivalence between the two research traditions. Rather, it identifies areas of functional overlap alongside substantial differences and therefore supports selective rather than comprehensive integration.

Multimedia-learning principles provide relatively well-defined and observable design categories that can be identified consistently across instructional materials and used for systematic coding. Their coverage of relational communication is more limited: behaviours such as smiling and humour, although well established in instructional communication research, are not represented as distinct multimedia-learning principles. Instructional communication provides a richer vocabulary for relational processes, but its constructs range from specific behaviours to broader relational concepts and may overlap in their behavioural manifestations. They therefore do not constitute a single feature-level coding system for asynchronous instructional videos.

6.4 A conceptual mapping tool linking multimedia learning and instructional communication

To address this feature-level analytical gap, the present thesis maps selected observable video features onto broader rhetorical and relational communication functions. Multimedia-learning principles provide observable design categories where an established principle corresponds sufficiently closely to the coded feature, while teacher clarity and immediacy serve as complementary, non-exclusive lenses for interpreting the features' primary rhetorical and relational functions. Additional instructional communication features are retained where no established multimedia-learning principle provides an adequate category.

In the analysed video, the coded features were organised into two broad groups according to their primary communicative orientation. The classification indicates predominant rather than exclusive functions; individual features may support both rhetorical and relational processes. Three coded features require clarification because their relationship to established multimedia-learning principles is less direct. Humour and smiling are treated as distinct instructional communication features because both are central to instructional immediacy research but neither is explicitly represented as a separate principle in Mayer's current multimedia-learning framework. Humour may overlap functionally with emotional

design, while smiling may be associated with embodiment; they are nevertheless retained as distinct features to avoid implying theoretical equivalence. Storytelling is also retained as an additional instructional communication feature because of its established relevance to attention, affect, and learning and because it constitutes a prominent feature of the analysed video.

The first group is classified as primarily relational and immediacy-related and includes personalisation, instructor face, humour, smiling, and storytelling. These features primarily communicate instructor presence, approachability, audience orientation, relational closeness, or emotional interest. Their classification does not imply that their effects are exclusively affective; several can also support attention, comprehension, and meaning-making.

The second group is classified as primarily rhetorical and clarity-related and includes signalling, visual elements, coherence, avoiding redundancy, segmenting, and examples. These features primarily guide attention, organise the instructional message, reduce unnecessary processing demands, and facilitate comprehension. Their classification likewise does not exclude relational functions.

Table 7 presents the conceptual mapping of the coded instructional-video features according to their primary rhetorical or relational communicative orientation and their relationship to teacher clarity and immediacy. The full operational coding scheme used in the empirical analysis is reported in Table 12.

Table 7: Conceptual mapping of multimedia-learning principles and coded instructional-video features onto rhetorical and relational communication functions

Primary communicative orientation	Design principle / coded feature	Description / theoretical basis	Instructional communication link and rationale
Primarily relational / immediacy-related	Embodiment (instructor face)	The embodiment principle concerns visible human-like gestures and movements in multimedia instruction. In the present study, instructor visibility was operationalised through segments in which the presenter's face appeared on screen and was interpreted as an embodiment-related feature.	Primarily related to nonverbal immediacy. Visible embodied social cues can increase perceived instructor presence, while nonverbal immediacy can reduce psychological distance between instructor and learner (Dixson et al. 2017; Fiorella 2021).
Primarily relational / immediacy-related	Humour	Humour is not a separate multimedia-learning principle in Mayer's current framework. It has partial functional overlap with	Related to instructional humour and immediacy. Appropriate humour may support positive affect, emotional interest, and approach orientation

Primary communicative orientation	Design principle / coded feature	Description / theoretical basis	Instructional communication link and rationale
		emotional design because both may evoke affective responses supportive of learning.	and may contribute to a more relational instructional experience (Mazer 2013; Plass and Hovey 2021).
Primarily relational / immediacy-related	Smile	Smiling is retained as a separate instructional communication feature rather than treated as an established multimedia-learning principle. Its relationship to emotional design or embodiment is indirect.	A nonverbal immediacy cue that can communicate warmth, approachability, and relational openness (Andersen 1978; Plass and Hovey 2021).
Primarily relational / immediacy-related	Personalisation	The personalisation principle concerns conversational forms of communication, including direct address and first- or second-person language.	Related to immediacy and affinity-seeking. Conversational language addresses learners as communication partners and may activate social responses, thereby supporting relational closeness (Mayer and Fiorella 2021b).
Primarily relational / immediacy-related	Storytelling	Storytelling is not included among Mayer's multimedia-learning principles. It is retained as an additional instructional communication feature because of its established relevance to attention, affect, and learning (Kromka and Goodboy 2019).	Related primarily to relational communication, but also to clarity and meaning-making. Narratives may create emotional interest and connection with the instructor while also making abstract content more concrete and meaningful (Kromka and Goodboy 2019; Mazer 2013).
Primarily rhetorical / clarity-related	Signalling	The signalling principle proposes that learners benefit when cues highlight the organisation and key information in instructional material.	Primarily related to teacher clarity. Highlighting relevant information directs attention and makes the organisation of the instructional message explicit (Mayer and Fiorella 2021b; Mazer 2022; van Gog 2021).
Primarily rhetorical / clarity-related	Multimedia visual elements	The multimedia principle proposes that learning can be supported through coordinated verbal and pictorial representations rather than words alone.	Related to message clarity and comprehensibility. Relevant visual representations can make the instructional message more accessible and support understanding (Mayer 2021; Titsworth and Mazer 2016)
Primarily rhetorical / clarity-related	Avoiding redundancy	The redundancy principle concerns avoiding unnecessary presentation of the same information in competing formats.	Primarily related to teacher clarity. Removing unnecessary duplication reduces competing message forms and can make explanations easier to process and follow (Mayer and Fiorella 2021b; Schrader et al. 2021).
Primarily rhetorical / clarity-related	Coherence	The coherence principle concerns excluding extraneous words, images, and other material that does not support the instructional goal.	Primarily related to teacher clarity. Excluding irrelevant material preserves message focus, reduces ambiguity, and supports comprehensibility (Mayer and Fiorella 2021b; Mazer 2022).
Primarily rhetorical / clarity-related	Segmenting	The segmenting principle concerns presenting multimedia material in manageable units rather than as one uninterrupted stream.	Related to clarity and message structure. Dividing an explanation into manageable units can make its organisation and pacing easier to follow (Mayer and Fiorella 2021; Titsworth and Mazer 2016).
Primarily rhetorical / clarity-related	Examples	Examples support the initial learning and understanding of concepts or skills by illustrating abstract information through concrete cases.	Related to teacher clarity and content relevance. Relevant examples can make abstract material more concrete, understandable, and

Primary communicative orientation	Design principle / coded feature	Description / theoretical basis	Instructional communication link and rationale
			meaningful (Titsworth and Mazer 2016; Mazer 2022).

Note. Author's own conceptual synthesis based on Houser and Hosek (2022), Mayer and Fiorella (2021b), Mazer (2022), and Mottet and Beebe (2006), with feature-specific sources indicated in the table. Features are classified according to their primary rather than exclusive communicative function. The full operational coding scheme used in the empirical analysis is reported in Table 12

The distinction indicates the primary rather than exclusive communicative function of each feature. Individual features may serve both rhetorical and relational functions.

7 MEASUREMENT OF AFFECTIVE RESPONSES

The purpose of the present chapter is to explore the methods designed to measure the level of engagement. Such measurements exhibit a remarkable degree of diversity, stemming from the fact that the concept of engagement and its indicators have been defined across a wide variety of models. Before reviewing the measurement methods, it was therefore necessary to delineate the concept of engagement from related concepts such as motivation, academic affect, and disengagement.

7.1 Detecting and measuring emotions and engagement

Though engagement is a complex construct, its measurement may have various benefits: the results can contribute to the identification of the contextual factors of education (such as classroom work organisation) that have the largest effect on academic outcomes, participation in a given context. Data collected on engagement can support national, institutional, or classroom-level prevention and intervention strategies by informing schools and instructors about the effects of different interventions, enabling them to identify the most appropriate ones for the given student population and adjust them to their actual needs. In the past decade, the types, thoroughness, and application of instruments measuring engagement have expanded to better inform researchers and decision-makers. The findings from these measurements indicate that students' engagement is significantly influenced by the emotional, social, and physical environments they encounter. This underscores the crucial role of instructional, institutional and system-level interventions in fostering individual engagement (Betts 2022).

Several studies, especially in the first phase of engagement research, approached engagement and disengagement as the *opposite ends of the same continuum* (see also Section 2.3.2). Some of the studies conceptualising engagement in this way defined different levels of engagement based on the types of emotions identified from facial expressions and the intensity and frequency of certain learning-related behaviours. Some studies defined *two* (Jang, Kim and Reeve 2016; Watson and Tellegen 1985a), others set *three* (Li et al. 2016), while others reported on *four* (D'Mello and Graesser 2011; Whitehill et al. 2014) or *five* levels of engagement (Altuwairqi et al. 2021; Schlechty 2011).

7.2 The relationship between achievement emotions and student engagement

A logical question may emerge during the reading of the current chapter: Can the results of emotional engagement be considered as reliable signs of engagement? Is there a consistent connection between the two? A comprehensive review of the relevant literature reveals that engagement is a broad meta-construct that attempts to describe a dynamically fluctuating attitude inferred from indicators such as behaviours, data, or student emotions. The results of the past 2-3 decades of multimodal measurements have demonstrated that emotional states, especially enjoyment, interest, and boredom, are significant and reliable yet not exclusive indicators of engagement.

The main reasons for discussing the measurement methods of academic emotions and learning engagement together are as follows:

(a) As detailed in Section 3.3, there is a close and reciprocal relationship between academic emotions and student engagement (Mazer 2017; Pekrun and Linnenbrink-Garcia 2022), and *achievement emotions relate to performance*, over and *above the influence of gender*, *age*, academic *ability*, and personality *traits*.' (Pekrun et al., 2023, p. 172, *emphasis added*).

(b) The strong relationship between engagement and emotions revealed by a large body of research in the past 30 years, and the lessons learned during the pandemic, drove the attention to the crucial importance of academic *emotions*, and the *emotional* turn emerging in several fields of study (such as Sociology, Political Theory and Cultural Studies, etc., Delahunty and Crawford 2024), underlying the significance of emotions in measuring engagement.

7.3 Signs of engagement and disengagement

As mentioned in Section 4.1, engagement can be identified and measured by several indicators. Some of these, mostly related to behavioural engagement, are reflected in *academic and behavioural data* on attendance, grades, homework, or course completion (Mandernach, 2015, as cited by D'Mello, Dieterle and Duckworth (2017). In the case of online learning activities, data of *student interactions*, such as log files recording the time and frequency of learning activities, number of pages read or edits made (Fredricks et al.,

2019, as cited by Fredricks 2022), can provide indications of the level of cognitive and behavioural engagement.

A scope of metacognitive *reflections, feelings*, and *emotions* perceived and accessible only to the individual (such as motivation, interest, subject-specific or general engagement, perceived attentiveness and dedication, attachment to teachers or peers) is a valuable source of insights, providing information on the level of cognitive and affective engagement.

The third group of indicators includes *bodily changes* and actions during the learning process: (a) physiological changes such as heart rate or skin conductance, body movements, especially facial expression, eye movements and posture. Some of these changes, and the emotional-mental states behind them (such as mind wandering, boredom, or flow), occur at a low level of awareness and can be recorded by human observers or machine sensors. (b) A scope of nonverbal signs and behaviours, such as facial expressions, eye and body movements, is relatively unlikely to be strategically controlled; thus, they can be useful indicators of individuals' internal emotional and cognitive states (Lewinski 2015b), especially low-conscious emotional states, such as boredom.

Boredom can be indicated by various indicators (D'Mello, 2007; Finkielsztein, 2020:7 and Ohlmeier et al., 2018), including

(a) *behaviours*, such as yawning, being late or engaging in irrelevant activities;

(b) *eye movements*, especially partial lowering of the eyelid, blank or wandering gaze, or increased rate of the closed state of the eyes;

(c) *body posture*, such as a relaxed, back-leaning posture, resting head on hand or fidgeting;

(d) *facial actions*, particularly AU4 (lid tightener), AU7 (brow lowerer), AU12 (lip corner puller, McDaniel et al. 2007).

Table 8 summarises the association between elements of (primarily non-intentional) nonverbal behaviour and facial actions (Action Units, AUs), as concluded by several

empirical studies that examined engagement using behaviour observation. The identification of engagement can be further facilitated by identifying behaviours associated with disengagement; the second half of the table, therefore, presents emotions and non-verbal behaviours associated with disengagement, based on the empirical studies listed in the table's footnotes.

Table 8: Nonverbal signs of engagement and disengagement

Type of engagement	Emotions experienced	Nonverbal behaviour elements and facial actions (Action Units, AU)
Engagement[1]	Neutral expression interest enjoyment positive valence increased arousal frustration, confusion	• Lid tightening (AU7)[3] • Eyebrow frowning (AU4) • Lip biting or sucking (AU 28) • Dimpler (AU14) – mental effort • Raising the upper lip (AU10) • Leaning forward (leaning back after the task) • Low intensity or lack of movement • Sitting in the middle of the chair
Disengagement[2]	Anxiety (fear) boredom negative valence high arousal confusion frustration	• Raising the outer edge of the eyebrow (AU2) • Frowning: (AU4) - frustration, mental effort • Pulled kip corner (AU12) • Eyebrow frowning (AU4) • Lip tightening (AU7) • smile • lack of activity • relaxed posture, leaning back or large movements • sitting on the edge of the chair, • fidgeting • leaning on the table • head turned away from the screen (sideways or upwards) • raising inner eyebrow (AU1) • frequent blinking and low heart rate - boredom

[1] Dael, 2013; D'Mello et al, 2007; Dragon et al 2008; Grafsgaard 2014; Ketonen 2017

[2] Almohammadi et al 2017; Ashwin, 2018; Bian et al, 2019; Bosch, 2015; D'Mello et al, 2007; 2009, 2017; Dragon et al 2008; Emotiw, 2018; Grafsgaard, 2013; Ketonen 2017; Whitehill et al, 2014, D'Mello, 2007; Finkielsztein, 2020; Ohlmeier et al., 2018; McDaniel et al., 2007

[3] AU: Action Unit, i.e., facial actions

As the list of action units in the third column of the table shows, some of the expressions and behaviours (eyebrow frowning and lid tightening) can be part of expressions indicating both engagement and disengagement. This can be explained by various different factors: engagement and disengagement can occur and emerge simultaneously, or because the same emotion can contribute both to engagement and disengagement, depending on length and context (such as smile, frustration, confusion (Arguel et al. 2019; D'Mello et al. 2014),

see Table 2, but most importantly because "*there is not necessarily a one-to-one correspondence between the occurrence of an emotion and a prototypical expression (see Ekman 1992). Rather, emotions are expressed in prototypical multimodal patterns of behaviors with striking variations* (Keltner et al. 2019, pp. 139–146). This conclusion reminds us of that examination of facial expressions and non-verbal behaviour can only be used to *infer* unobservable mental and emotional states (D'Mello, 2017).

7.4 Methods applied to measure engagement

The identification and measurement of emotions was hindered by several factors until the last decades of the 20th century, such as (1) the lack of adequate conceptualisation of emotions, (2) the ethical and methodological dilemmas related to triggering strong emotions in a laboratory environment (Plutchik, 1982), and (3) uncertainty regarding the interpretation and analysis of the data collected using automated and real-time measurements. Since then, several conceptual frameworks have been developed and refined based on empirical studies to support the definition, comparison, and classification of emotions (Plutchik, 2001; Ekman, 2003; Russell, 1983; Ekman and Cordaro, 2011; see also Section 3.2). In the last 20 years, most empirical studies in the field of emotions and student engagement applied, elaborated and tested various methods to measure emotions and attitudes using a wide range of methods, including human coding, self-reports, retrospective interviews, computer-based engagement detection from facial expressions, body or gaze movements, physiological, administrative, or log file data (see Table 9).

In the case of boredom, measurement is particularly challenging due to its subjective, complex, and temporally unstable nature (Mills et al. 2018). Its semi-conscious nature reduces the reliability of self-reported measures (Eastwood et al. 2012). Theoretical papers and meta-analyses (Chrysafiadi and Virvou 2013; D'Mello and Kory 2012; Martarelli and Jost 2024) therefore recommend combining self-reports with measurements that provide information via physiological, visual, or vocal channels to enable the recognition and deeper understanding of hidden emotions. The advantage of automated psychoneurophysiological approaches is that they enable systematic monitoring of fluctuations in boredom over time (Wolff & Martarelli, 2020).

The current summary outlines an overview of the methods applied to measure engagement and emotions in the past two decades, relying on seven methodology reviews (D'Mello,

Dieterle and Duckworth 2017; Dewan, Murshed and Lin 2019; Yadegaridehkordi *et al.* 2019; Dubbaka and Gopalan 2020; Betts 2022; Fredricks 2022; Martarelli and Jost 2024).

Measurement methods can be categorised based on their different features:

- Considering the channel of information, these measurements can be *textual, visual, vocal, physiological, and multimodal* (Yadegaridehkordi et al. 2019). Mixed-method and multimodal studies provide opportunities to triangulate results across different data sources, thereby increasing objectivity.
- According to the level and the type of *the participants' involvement* in the engagement detection process (Dewan, Murshed and Lin 2019), it can be *automatic* (every type of automated measurements), *semi-automatic* (observations with semi-automated data coding) or *manual* (observations with manual coding and self-reports). The "automatic" category is also referred to as AAA (Advanced, Analytic, Automated) measurements (D'Mello, Dieterle and Duckworth 2017).
- Based on *the nature of the data source*, (1) *subjective* methods provide information on subjective thoughts and feelings that can be accessed only by the participant, and can be obtained through their self-reports, while (2) *objective* measures rely on externally observable indicators such as behavioural, physiological, or neurological measures (Martanelli & Jost, 2024).

When applied specifically to boredom, this distinction reveals an important pattern. Despite the limitations of self-reports and the intensive development of affective computing and other computer-vision-based tools, (Yadegaridehkordi, 2019; Merrifield & Danckert, 2014; Kim et al., 2018), academic boredom is still predominantly identified and measured by text-based self-report scales, though the usage of objective measures is increasing (Martarelli and Jost 2024).

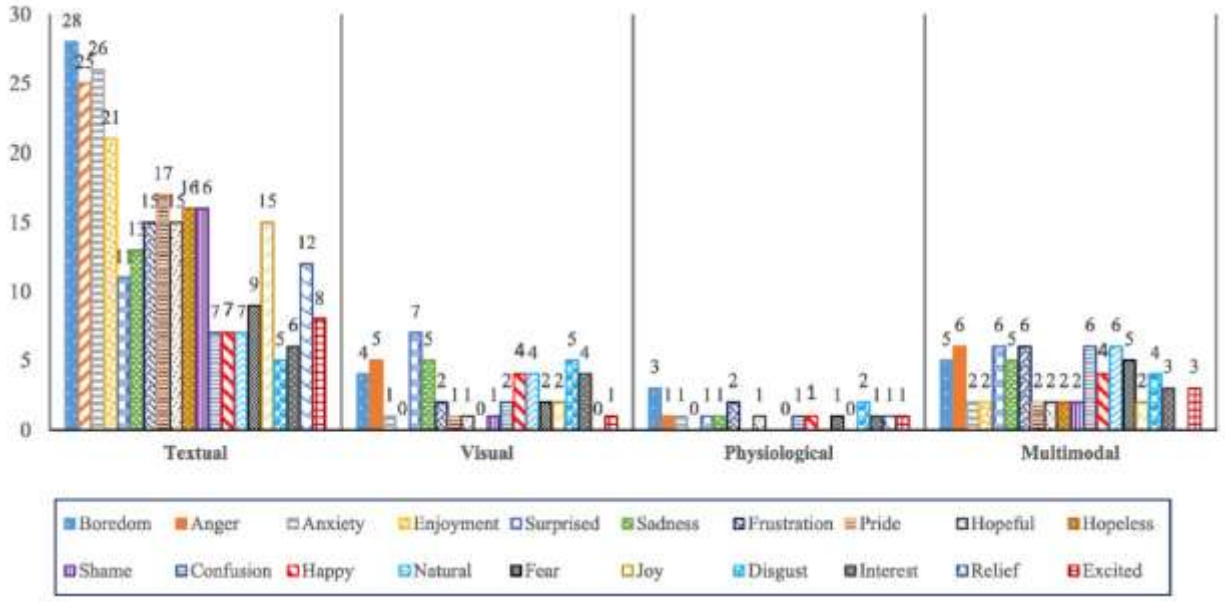


Figure 9: Distribution of emotional states by affective measurement channels.
Source: Yadegaridehkordi *et al.* 2019, p. 11

Table 9 summarises the main measure types and their features, in line with the categories and aspects outlined in the above-mentioned reviews.

Table 9: Categorisation of measurement methods for learning engagement

Classification aspect	Subjective measures		Objective measures
Typical methods	- self-report: questionnaire, - interview, - self-reported text - experience sampling - teacher ratings	observational techniques applied by external coders: - rating scales of behaviour - qualitative studies	Sensor-free[2] data engagement tracing - log file analysis - administrative data
			Sensor-light data[3] - facial expression analysis - body movement analysis: head pose, eye closure, gaze direction - keystroke dynamics - speech - prosody and intonation - seat back and backrest pressure
			Sensor-heavy data[4] - EMG, EKG, ECG, heart rate, EOG, EEG, skin conductance, electrodermal activity, blood volume pressure
Level of awareness (D'Mello, Dieterle and Duckworth 2017; Fredricks 2022)	High: explicit: - metacognitive reflections, - subjective feelings, - thoughts	Low: implicit - Annotations of visible behaviours - by human observers	Low: implicit - Neurobiological changes measured by machine sensors - inferring emotional states based on computer vision-based tools
Possible measurement channels (Yadegaridehkordi et al. 2019)	textual	textual	- visual - physiological - vocal

[2, 3, 4] (D'Mello et al., 2017)
Author's own compilation based on Dewan et al., 2019; D'Mello et al., 2017; Fredricks, 2022; Martarelli & Jost, 2024; Yadegaridehkordi et al., 2019

Sections 7.4.1 and 7.4.2 provide a brief overview of the advantages and limitations of the different measures.

7.4.1 Subjective methods

Subjective methods include both self-reported measures and manual coding taken by external observers or teachers.

Self-reports most often take the form of standardised questionnaires but can also be implemented as any other kind of oral or written textual report shared by a participant, during a certain period (e.g., experience sampling), or before or after the process to be investigated, e.g., such as day reconstruction or interviews (Landmann 2023).

The most important quality of self-reports is that (a) they do not rely on inferences (as in the case of objective measures) but provide insight into the participants' perceived emotions and experiences; therefore, (b) this type of measure is considered to be the most *reliable* (Appleton et al., 2008, as cited in Fredricks, 2022) (c) and the *most applied* (D'Mello, Dieterle and Duckworth 2017) strategy in the context of learning. Further advantages of this measurement type include (d) ease of data collection, (e) suitability for large numbers of participants, (f) time- and cost-effectiveness, and (g) good comparability of results.

Numerous general and domain-specific psychometrically validated scales inform the development of self-report questionnaires for measuring self-perceived emotional states. Focusing specifically on boredom, there are currently approximately 10–12 standardised English-language self-report measures of boredom (Martarelli and Jost 2024). Self-report questionnaires, however, provide only limited access to low-awareness states such as boredom, confusion, or frustration (Fredricks 2022).

Human observers (trained observers or teachers) usually report on participants' behaviour through informant questionnaires, manually or digitally completed observational checklists, or less-structured written reflections. Though observers can rely only on the visible behaviours of participants, based on which they can infer their engagement or emotions, an observation study conducted during e-learning activities found that observers of students' nonverbal behaviour predicted students' test performance with the same accuracy as the students' pre-test scores (Whitehill et al. 2014).

A review examining measures across their channels between 2010 and 2017 found that the majority of engagement measures were text-based: they analysed self-reports collected via questionnaires or textual reports (Yadegaridehkordi et al. 2019). These methods, however, have multiple limitations: (a) they usually provide a general picture of participant engagement throughout the learning phase. Given that engagement is a fluctuating, dynamic phenomenon, subjective methods are limited in their inability to track these changes in a nuanced way. Further limitations include (b) the lack of exact frames of reference in the case of rating their own emotions and attitude, (c) memory recall, and (d) social desirability bias. Manually coded observations are time- and work-intensive, obtrusive, and can be biased by the halo effect or by the heterogeneity of the frames of reference applied when evaluating a person or action.

7.4.2 Objective methods

7.4.2.1 Sensor-free data analysis

The analysis of administrative data on student attendance and performance, as well as log files recording student interactions during online learning activities, does not require additional data collection, sensors, or special devices; it is automatically collected. These data are typically stored in different databases and thus can be regarded as automatically collected, sensor-free data (D'Mello, Dieterle and Duckworth 2017). The relationship between administrative or log file data and student engagement is nevertheless based on assumptions; therefore, some taxonomies label this data type as semi-automated (Dewan, Murshed and Lin 2019).

7.4.2.2 Automated measures

Methods applying automated data collection are based on the fact that "cognitive and affective states reflecting different components of engagement are associated with responses at multiple levels (neurobiological, physiological, bodily expressions, overt actions, metacognitive, and subjective), some of which are detectable by machine sensors" (D'Mello, Dieterle and Duckworth 2017, p. 107). Automated measurements can also be grouped according to the type of involvement they are best suited to identify:

- Changes in central physiology and eye gaze movements may be the most suitable for measuring *cognitive engagement* (Berka et al., 2007; Deubel & Schneider, 1996; Hoffman & Subramaniam, 1995; Rayner, 1998 as cited by D'Mello, Dieterle and Duckworth 2017).
- Facial actions and peripheral physiology provide information on *affective engagement* Matsumoto et al., 2008)
- Behaviours, log file analysis, and administrative data may indicate levels of *behavioural engagement* (Baker & Ocumpaugh, 2015; Baker & Rossi, 2013; Bulger, Mayer, Almeroth, & Blau, 2008; Gobert et al., 2015, as cited by D'Mello, Dieterle and Duckworth 2017).

Automated data collection can be implemented when applying (sensor-heavy) physiological measurements or in the case of (sensor-light) computer-vision-based measures (D'Mello, Dieterle and Duckworth 2017).

Sensor-heavy neurophysiological measures

These measures are also referred to as sensor-heavy methods, due to the specific equipment with limited accessibility required by them, as well as the high level of analytical expertise required to interpret these data.

Computer-vision-based measures

Computer-vision-based measures are unobtrusive, easily quantifiable, rich in data, cost- and effort-effective, easy to administer, and generally reliable. Due to intensive technological developments and the increasing accessibility, accuracy, and ease of use of software and devices enabling various real-time or objective measurements, the rate and role of these types of measurements are increasing. They enable the collection of not only static but also dynamic data reflecting changes in engagement over time, while causing minimal disruption and human workload. The large amount of data obtained in this way "allows for within-person, repeated measures analytic methods to be applied when engaging in research" (Betts 2022, p. 652).

7.5 Automated facial expression analysis

The relationship between facial actions and basic emotions (Ekman, 1978) is relatively widely studied and defined (Dewan, 2019; Sass & Vinczéné Fekete, 2022). The current study applied FaceReader 9.1 (Noldus International Ltd) as a measurement tool for facial expression analysis that combines algorithmic face modelling, expression classification, and FACS-based action-unit coding developed by Ekman and Friesen (1978), enabling unobtrusive webcam-based recording and the modelling of emotions as dynamic processes (Teixeira et al., 2012, as cited by Ladmann, 2023). Besides FaceReader, several similar applications are used for computer-vision-based facial expression analysis, such as NViso, iMotions, OpenFace, or AFAR. Recent validation work suggests FaceReader's basic-emotion recognition can match or outperform human coders (Lewinski, 2015a; Hetland et al., 2019), including evidence of 88% accuracy for FaceReader versus 85% for humans in standardised databases (Lewinski et al., 2014).

The reliability of emotion identification varies: happy and neutral show 96% and 94%, respectively, while anger shows 76% and sadness shows 86%. Emotional states that are expressed less universally, such as boredom, can be identified with lower reliability, and

in this case, the process relies on analysing the frames within a time window of more seconds. Boredom, the affective state essential in the context of the current study, requires a 5-second time window for identification; however, it is detected with high precision but relatively low sensitivity. According to a validation study (Gudi 2017), more than 84% of cases labelled as boredom were indeed cases of boredom. At the same time, the system recognised only 55% of all boredom cases, indicating that boredom might be frequently overlooked: 20% of the true boredom cases were classified as 'none' rather than boredom, in line with earlier empirical results (McDaniel et al., 2007; Dewan, 2019; Dubakka, 2000).

The current chapter, therefore, provides deeper insight into the methodology it applies to support understanding of the current research results. FaceReader 9.1 has been trained to identify 6 emotions (happy, sad, angry, surprised, scared, disgusted, identified as basic or universal emotions by Ekman, (1970), and additionally neutral expression, contempt and facial states of head and eye movements. A relatively new feature of the software is the ability to identify and measure more complex emotional states, such as boredom, frustration, and interest. The software infers basic emotions and additional states (e.g., neutral, contempt, eye/mouth movements) from facial actions (Ekman, 1970). The analysis relies on faces via automated face finding, modelling 468 facial key points (Loijens and Krips 2025), and classifying expressions using deep neural networks and FACS-based action units (Ekman & Friesen, 1978). It calculates estimated arousal and valence, and secondary affects such as confusion, interest, and boredom for videos, based on combinations of facial muscle movements and their intensities, per frame or over 2-5 second time windows (Gudi, 2017).

Data derived from automated facial expression analysis should be interpreted with methodological caution, as the underlying computer vision and affect detection techniques are still evolving and their analytical application is not yet fully standardised. Although the use of FaceReader and similar tools is growing, the limited number of reliable, widely used analytical strategies and the variation across studies indicate that this field remains methodologically experimental. It follows that an important distinction is maintained throughout the present thesis between facial actions, automatically generated classifications, and experienced emotions. Facial movements constitute the observable input to the analysis, while emotion labels constitute outputs generated through a

classification model. The latter should therefore not be interpreted as direct readings of an individual's internal emotional state. Barrett et al. (2019) show that facial expressions do not consistently correspond to specific emotions, while D'Mello, Dieterle, and Duckworth (2017) similarly note that bodily signals can be difficult to interpret as specific mental states. Accordingly, the main evidential value of automated facial analysis in this study lies in identifying temporal variation, peaks, recurring patterns, and differences associated with instructional conditions. The interpretation of the psychological meaning of those patterns additionally draws on self-report, interview data, and theoretical context.

7.6 Methodological implications of multimodal measurements

The rationale for multimodal measurement in the present study extends beyond the expectation that different measures should confirm one another. Different methods provide access to partly different aspects and temporal levels of learners' responses, and discrepancies between them may therefore be theoretically informative. Greene, Caracelli and Graham (1989) distinguish several purposes for combining methods, including triangulation, complementarity and initiation. The present research draws particularly on triangulation and complementarity: convergence across methods can strengthen an interpretation, while divergence can clarify the boundaries of particular measures or reveal differences between experienced, expressed, remembered, and automatically classified responses.

Bodily signals only allow us to *infer* mental and emotional states. These signals can be coded by human observers or computers. Automated coding can offer advantages over human observation in consistency, temporal precision, scalability, and the ability to process large datasets. However the lesson concluded from the review of studies applying multimodal measurements underscore that even this “the inference process is rife with ambiguity at multiple levels” (D’Mello, 2017, p. 117), due to several reasons: the effect size of the correlation between self-reported and automated detection is usually weak or medium; bodily signals of the same emotion may be different in the case of the same individual; due to the potential for errors arising from the fact that the machine learning process involves human coding (D’Mello et al., 2017).

The above conclusions suggest two important methodological considerations for research.

a) Different measurement methods provide access to different types of emotional and mental states: while *eye gaze and changes in central physiology are related to cognitive engagement*, *facial actions* and changes in *peripheral physiology* indicate changes in *emotional* engagement. *Analysis of log file analysis and behaviours* provides information on *behavioural* engagement (Baker & Ocumpaugh, 2015; Keltner & Ekman, 2000; Marshall, 2005; Matsumoto et al., 2008; cited by D'Mello et al., 2017). It is therefore necessary to employ multimodal measurements to facilitate a nuanced examination of the features and factors influencing engagement.

b) As recommended by the same study, the adequate analysis and interpretation of data from automated measures requires the collaboration of interdisciplinary teams, including education and computer science researchers (D'Mello et al., 2017).

8 METHODOLOGY

The current chapter begins by exploring theoretical and empirical studies that address the challenges and opportunities of measuring student engagement, including the methods used, their advantages, and potential limitations. Section 8.1 presents the research design framework established for the current study, based on the conclusions drawn from the methodological overview. As the study's primary measurement tool was software that estimates emotions based on facial expression analysis, to facilitate comprehension of the dilemmas related to these measurements, a separate section (8.7.3) provides insight into the software's operation and the technical background of the data calculations.

8.1 Framework of research design

The methods applied in the current study were informed by the findings of several reviews on the detection of learning-related engagement and emotion (Dewan et al., 2019; D'Mello et al., 2017; Dubbaka & Gopalan, 2020; Yadegaridehkordi et al., 2019). Since, in learning contexts, each method may be affected by potential sources of bias, several studies (Mazer & Graham, 2019; Yadegaridehkordi et al., 2019) recommend a multimodal approach, for example by incorporating data from different channels and drawing on both human-assisted and automated detection (D'Mello & Kory, 2012), or by combining the results of facial analysis with those of physiological measurements (Zhong et al. 2017). The current study therefore adopted a multimodal measurement approach combining automated facial expression analysis, questionnaire data, and semi-structured interviews. These methods differ in measurement channel, temporal resolution, degree of conscious access, and inferential distance from learners' experiences. Their combination was intended not only to examine convergence across indicators but also to identify complementary and potentially divergent aspects of students' affective responses (Christensen, Brewer and Hutchinson 2018; Jewitt 2014; Poria 2016).

The research design was informed by Emotional Response Theory (ERT; Mottet, Frymier and Beebe 2006), which provides a theoretical reason to examine emotional responses in relation to instructional communication and subsequent approach-avoidance tendencies. ERT is used here as an explanatory framework rather than as an assumed causal model. Subsequent testing has provided only partial support for its proposed causal sequence (Horan, Martin and Weber 2012). Accordingly, the present design examines whether

communicative and design features co-occur with theoretically relevant patterns of learners' affective responses and does not interpret temporal precedence alone as evidence of causation.

The multimodal design additionally reflects the assumption that different measurement channels provide non-equivalent evidence about learners' responses. Self-report and interview data provide access to learners' consciously interpreted and retrospectively articulated experiences, whereas automated facial analysis provides temporally fine-grained indicators of facial activity. The purpose of combining these sources is therefore both triangulation and complementarity (Greene, Caracelli and Graham 1989).

The current thesis applied a strategy recommended by D'Mello and his colleagues (2017) as follows: the first step is to measure an emotional state that indicates learning (e. g., boredom), the second step is to observe its possible triggers (e.g., the events preceding its appearance). This can lead to the modelling of the possible triggers or antecedents of the given emotion (see Figure 10). In line with the ERT perspective, this strategy enabled the examination of emotional responses during learning, alongside the communicational and design-related features that may have preceded them.

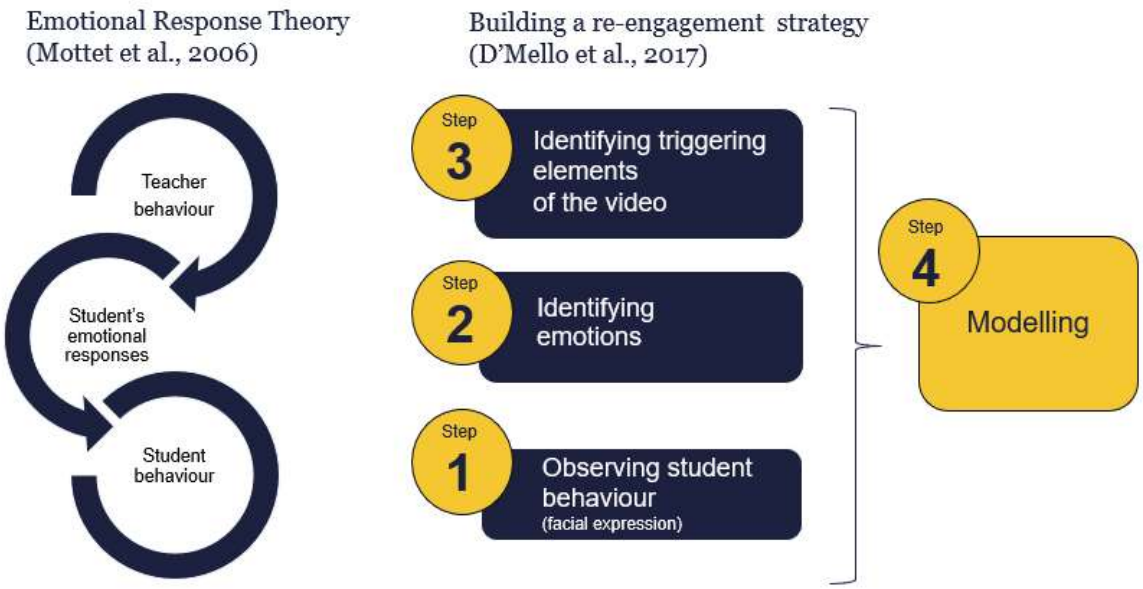


Figure 10: Conceptual framework and analytical steps for detecting emotional responses and modelling re-engagement strategies.

(Author's own compilation, based on Mottet et al. 2006 and D'Mello et al. 2017, pp.118)

Though the presence and effects of multimedia principles have increasingly been studied (Çeken & Taşkın, 2022; La Torre & Désiron, 2024) and a large number of studies applied multimodal measurement of emotions during learning (Whitehill *et al.* 2014; D'Mello, Dieterle and Duckworth 2017; Yadegaridehkordi *et al.* 2019; Noroozi *et al.* 2020), to the best of our knowledge, no empirical studies have investigated the relationship between automatically detected facial expressions and the applied multimedia learning principles, as presented by the strategy presented by Figure 10.

8.2 Sample

The sample included 55 Communication Science BA students (male = 13, female = 42) aged 19 to 23 from Corvinus University of Budapest. A non-probability convenience sampling strategy was used within a predefined course-based population. Students enrolled in the course taught by the instructor appearing in the instructional video were invited to participate voluntarily as an optional course activity.

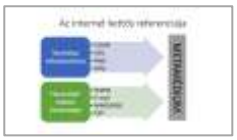
Data were collected in three waves over a 2.5-year period (in 2021, 2022, and 2024) from students enrolled in the same course across different semesters. The protocol, instruments, and analytic procedures were held constant across waves. No substantive differences in the main outcome patterns were observed across waves; therefore, the data were pooled. After data cleaning (excluding recordings affected by technical issues), the final dataset comprised data from 44 interviews, 51 questionnaires, and 50 video recordings. The procedures for data collection and storage were implemented in accordance with the ethical considerations outlined in Section 8.6, taking into account the sensitivity of the facial videos and the participants' vulnerable status.

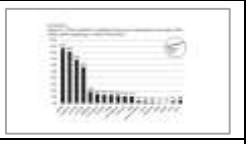
Data were also collected during additional instructional video viewing sessions, involving 45 additional students. These datasets were not included in the present analyses because the respective learning situations differed substantially in several potentially relevant contextual characteristics, including video length and topic, participants' age, and whether the associated course was compulsory or optional. Given this heterogeneity, direct comparison of the datasets would have introduced additional sources of variation that could not be adequately accounted for within the scope of the present study. The current analyses were therefore restricted to the more homogeneous dataset described below, while the remaining data may provide a basis for future comparative research.

8.3 Stimulus

The stimulus viewed by students was a 22.4-minute instructional video produced by the course instructor, covering the topic of “Brand destruction” that related to their course curriculum. The language of the video and the data collection was Hungarian, the participants' native language. The video incorporated a range of instructional communication strategies and multimedia design principles intended to support students’ understanding, attention, and engagement. Table 10 presents a detailed, segment-level overview of the instructional video, including its structure, applied instructional strategies, and multimedia design features.

Table 10: Segment-level overview of the instructional video stimulus and its instructional strategies and multimedia features

Segment number	Time length (mm:ss)	Topic, method, tools	Applied multimedia principles	Slide preview
Segment 1	00:08	opening video, showing a paper bag being crumpled up & the video title	Multimedia	
Segment 2	00:10	video title greeting selfie video check-in Indicating video topic	Personalisation Smile Embodiment Smile	
Segment 3	00:45	slide with short text	Coherence	
Segment 4	1:32	slide with longer text and picture	Redundancy Multimedia Humour	
Segment 5	1:26	text structured in a flow chart familiar style guidance from the teacher to support students' metacognition	Lack of coherence, cognitive overload Personalisation	
Segment 6	8:37	text structured in a table, with on-screen highlights English language slide and Hungarian language spoken narration short examples and story-like references	Lack of coherence Lack of segmenting Example Storytelling Humour Signalling	
Segment 7	1:01	graph with on-screen annotations	Multimedia Signalling Segmenting Example	

Segment number	Time length (mm:ss)	Topic, method, tools	Applied multimedia principles	Slide preview
Segment 8	0:30	Static slide with bar chart and extraneous information	Multimedia Segmenting Lack of coherence	
Segment 9	1:30	graph with on-screen annotations and signals	Multimedia Signalling Segmenting Lack of coherence	
Segment 10	1:08	diagram with on-screen annotations and signals	Multimedia Redundancy Signalling	
Segment 11	0:32	graph with on-screen annotations signals	Multimedia Signalling Segmenting Lack of coherence	
Segment 12	1:57	diagram with on-screen annotations signals	Multimedia Signalling Segmenting Coherence	
Segment 13	0:05	short video-check-in while the teacher's face was simply seen during the explanation gossip-style storytelling	Storytelling Embodiment	
Segment 14	0:08	gossip-style storytelling	Humour Storytelling Multimedia Example Coherence Embodiment	
Segment 15	1:37	short video-check-in while the teacher's face was simply seen during the explanation gossip-style storytelling	Humour Storytelling Embodiment	
Segment 16	0:21	video message ending with a humorous twist	Humour Personalisation Embodiment Smile	

8.4 Measurement instruments

As recommended by boredom and academic emotion research scholars (see Chapter 7), the current study relied on multimodal data, including quantitative and qualitative data, that were varied according to their resources, channels, level of automation and level of involvement, to enable multimodal triangulation. Questionnaire and interview responses were used primarily to contextualise and interpret the facial-analysis indicators and to examine convergence and divergence across measurement channels rather than to function as interchangeable measures of the same latent state.

8.4.1 Questionnaire

The self-report questionnaire was compiled based on the following standardised scales:

a) the Positive-Negative Affectivity Scale (PANAS) (Gyollai et al., 2011; Watson, Clark and Tellegen 1988), to follow preceding emotional state bias;
b) the Student Interest Scale, evaluating cognitive and emotional interest (Mazer 2012)
c) and Verbal Teacher Immediacy Scale (Gorham 1988).

The questions, the information on the scoring procedures applied to the scales and the reliability table are available in Appendix D.

8.4.2 Interview

The semi-structured interviews explored which features made the video engaging and which parts were difficult to follow. Participants were also asked which instructional elements in the video, including text slides, graphs, tables, pictures, the teacher's face, and storytelling, most supported or hindered their attention. The interviews concluded with a brief introduction to the software methods and features, followed by a discussion of participants' viewing experiences and selected episodes during which FaceReader detected increases in relevant facial-expression indicators. The retrospective interviews were semi-structured and organised around two main questions, which are provided in Appendix H. While these core questions guided all interviews, their wording and sequence varied somewhat across participants, and follow-up questions were used where appropriate to clarify or elaborate on participants' responses.

In Section 8.7.1 describing the coding scheme and Chapter 9 summarising results, participants and the instructor featured in the video are identified using anonymised codes (P1, P2, etc. for participants and V for the instructor). Verbatim interview excerpts are reported in text as unpublished qualitative material.

8.4.3 Facial analysis

FaceReader 9.1 software applied for facial expression analysis infers emotions and cognitive-emotional states based on facial actions (Ekman, 1970). The analysis relies on faces via automated face detection, modelling, and classifying expressions using deep neural networks and FACS-based action units (Hjortsjö, 1970; Ekman & Friesen, 1978). It calculates estimated arousal and valence, and basic emotions (happy, sad, angry, scared, surprised, contempt) and secondary affects such as confusion, interest, and boredom for videos, based on combinations of facial muscle movements and their intensities, per frame or over 2-5 second time windows (Gudi, 2017).

8.5 Procedure

Data collection followed a five-step protocol (see Figure 11). Students (1) completed a baseline self-report questionnaire on their current emotional state; (2) watched the instructional video while their webcam face recording was captured; (3) completed the post-video questionnaire (including cognitive and emotional interest and perceived teacher immediacy); (4) took part in a semi-structured retrospective interview about segments and features that elicited high boredom or high interest; and (5) reviewed automatically generated facial-analysis outputs and commented on their feelings during peak-emotion moments. Sessions were conducted individually, in person, at Corvinus University of Budapest, and lasted approximately 45 minutes (interviews: 5–10 minutes, depending on participant verbosity). The study materials and all data collection were conducted in Hungarian, the participants' native language.

Each student was recorded in a one-on-one setting. During video viewing, participants were left alone in the room to minimise potential observer effects associated with the researcher's presence. Students watched their own instructor's video presentation while they were recorded by a web camera, as recommended by Whitehill and his colleagues (2014), with the difference that in this case, the web camera was built into the computer,

further increasing the unobtrusiveness of the observation. Observation bias was reduced by leaving the participants alone in the classroom during the video viewing. Both before and after watching the video, they completed an online self-report questionnaire, followed by a retrospective student interview, which provided an opportunity for students to give feedback on their perceived emotions compared with the facial analysis results (see Figure 11).



1. Questionnaire:

emotional state
before the video



2. Video watching



3. Questionnaire:

- emotional state after the video
- cognitive interest
- emotional interest
- perceived nonverbal teacher immediacy



4. Interview

- overall impression
- parts assessed as boring
- parts assessed as appealing/engaging

Figure 11: Procedure of the observation

8.6 Ethical considerations

The procedures for data collection and analysis have been developed based on careful consideration of the risks of different types of harm, their likelihood of occurrence, and the measures that can minimise these effects. Moreover, data collection was strategically designed to ensure that participation would be both beneficial and positive for participants, thus counterbalancing any potential stress or time expenditure (Iphofen 2011). Table C1 and

Table C2 in Appendix C provide a detailed analysis of the aspects, opportunities, risks, and benefits of participation, which are discussed as an important part of the research design.

As facial videos are sensitive data, their management and transparency regarding the automated tools used have been implemented in accordance with the European Code of Conduct (ALLEA - All European Academies 2023) and the questionnaires and forms provided by the Research Ethics Committee of Corvinus University of Budapest (Corvinus University of Budapest 2020). Compliance with the university regulations was supervised by the institution's data management officer.

A further sensitive aspect concerns potential power asymmetries between students and teaching staff, which can increase students' perceived vulnerability in research settings. To minimise the risk of perceived pressure, anxiety or negative consequences on student-teacher relationships, participants' anonymity was ensured by using participant codes instead of names, and teaching staff had no access to the raw database or detailed analyses, receiving only aggregated summaries. Further details can be found in Appendix A.

8.7 Data analysis

The data analysis methods were selected based on the findings of the methodological review presented in Chapter 7. The scope of this dissertation allows for a detailed presentation of the automated facial analysis data. The questionnaire and interview data can be examined here only to a more limited extent; therefore, they are presented in less detail and primarily used to support the interpretation of the facial analysis findings.

Table 11: Data collection and data analysis methods

Data source/instrument	What it captured	Related RQs	Analysis
Instructional video	video presentation	RQ1 RQ2	semi-automated coding, with a negotiated agreement approach
Automated facial expression analysis (FaceReader)	moment-to-moment affective indicators	RQ1 RQ2 RQ3	paired sample t-test, Wilcoxon tests, ANOVA, cluster analysis
Questionnaire	perceived affectivity, emotional and cognitive interest, teacher verbal immediacy	RQ2, RQ3	correlation analysis
Semi-structured interviews	retrospective self-reports and explanations	RQ1, RQ3	deductive qualitative coding

8.7.1 Coding-based qualitative content analysis of instructional video

The coding framework was conceptually aligned with the four-dimensional organisation presented in Figure 8, allowing the coded video features to be interpreted in relation to both rhetorical and relational communication functions. The instructional video used by the present research as a stimulus has been analysed according to the categories outlined in The Cambridge Handbook of Multimedia Learning (Mayer and Fiorella 2021). Table 12 defines the coding criteria for the video analysis, along with examples.

Table 12: Coding framework for the qualitative analysis of instructional video (1-second intervals)

Higher-order category	Design principle	Coded design feature (1/0)	Definition	Features of the video and quotes from the instructor
Immediacy	Embodiment	Face / No_face	Onscreen human(like) gestures and movements enhance learning	Appearance of the presenter's face on the screen
	Humour	Humour / no humour	"People learn better from multimedia materials when the materials' design induces emotional states conducive to learning without significantly increasing the need to process irrelevant information."[1]	Humorous expressions or actions, "I wish everyone a happy Easter and at least as many *delicious eggs as the Easter Bunny brings you. Oh, that was a product placement, damn it.*" Source: V
	Smile	smile[2] / no_smile		presenter's smile
	Personalization	Personal / No_persona-lization	Applying 2nd person talk; statement supporting students' metacognitive skills; conversational, empathetic and inclusive style.	"*Dear Colleagues, today we are going to review how it happens when …*" "*…how Pepsi and Coca-Cola kick each other around*" "*fortunately, this is quite a storytelling study*" Source: V
	Storytelling[3]	Story / No_story	Discourse elements that describe an important event or "symbolic actions	A story is used to illustrate the topic learned. "*So, here's just a bit about Dave Carol. He started out as a pretty unknown*

[1] Plass and Hovey 2021, p. 324

[2] This category is not among the multimedia design principles set by Mayer and Fiorella 2021

[3] This category is not among the multimedia design principles set by Mayer and Fiorella 2021

Higher-order category	Design principle	Coded design feature (1/0)	Definition	Features of the video and quotes from the instructor
			that have sequence and meaning for those who […] live, create and interpret them"[4]	*country musician in the US, when one time...* " Source: V
Clarity	Signalling	Signalling / No_signalling	People learn better when cues spotlight the key information.	Dynamic drawing, on-screen annotations; highlighting the essential parts of the text or the graphics
	Multimedia	Visual_element / No_visuals	Learning from words and pictures is more effective than from words alone	presenting verbal and pictorial information simultaneously
	Avoiding redundancy	Not_redundant / Redundant	People learn better when the same information is not presented in more than one format.	Oral verbal information is illustrated by visual elements.
	Coherence	Coherent / Not_coherent	Extraneous words and images are excluded.	The visual and oral messages include only the necessary information. There is no mismatch between the slide and the narration language.
	Segmenting	Slide_first_minute / Slide_rest	People learn better when a multimedia message is presented in learner-paced segments.	The video uses various tools to avoid monotony; the same slide is not displayed for more than 1-2 minutes, and new information is provided through animation.
	Examples	Example / No example	Examples help in the initial learning of cognitive skills.	Examples are used to illustrate the topic learned. "*Similarly, the competition between McDonald's and Burger King can also come to mind*." Source: V

Note. Author's own compilation based on (Houser & Hosek, 2022; Mayer & Fiorella, 2021b; Mazer, 2022; T. Mottet & Beebe, 2006) Quotations marked as **V** are verbatim excerpts from the instructor featured in the analysed instructional video and are presented as unpublished study material.

Coding unit and decision rules:

- Unit of coding: 1-second interval (binary per second: 0 = absent, 1 = present).
- Presence rule: a code is marked present in each second if the feature is visible/audible at any point in that second.
- Overlaps allowed: multiple features can be coded as 1 in the same second.

[4] Kellas 2008; Fisher, 1984, p. 2, cited by Kromka and Goodboy 2019

The coding framework included two communicational tools that are not explicitly included in the set of multimedia learning principles defined by Mayer and Fiorella (2021): (a) *storytelling* has been included and examined in the current study due to its evidence-based effectiveness in promoting students' attention, affect and performance (Kromka and Goodboy 2019, cited by Mazer 2022), on the one hand, and due to its prominent role in the video stimulus. *Smile* has been included because it is one of the few central immediacy behaviours (Andersen, 1978) that can be applied even in the case of instructional videos.

8.7.2 Self-reported experiences

Given the focus of the current thesis on multimodal indicators and their alignment with interview-based explanations, questionnaire findings are not reported in detail here and are reserved for separate reporting

The interview coding framework was developed jointly by two coders and refined through a negotiated agreement approach. After independently coding 10% of the interview transcripts, the authors compared and discussed coding differences, fine-tuned the framework accordingly, and the first author coded the remaining 90% in line with the agreed criteria (O'Connor and Joffe 2020).

The interview analysis followed a deductive qualitative approach. The coding frame (see Appendix D) was developed with predefined design elements, while the other part was generated inductively from the interview data through constant comparison (Glaser & Strauss, 1967). The current paper discusses the codes and results of the deductive approach, while the analysis of data identified through the inductive approach falls outside the scope of the present thesis and will be presented in a separate paper.

8.7.3 Facial expression data

The automated facial analysis generated a large, temporally fine-grained dataset. Across participants, the analysed recordings represented 72,600 seconds of video data, yielding approximately 1.09 million FaceReader data points for the facial expression and affective indicators included in the dataset. The volume and temporal resolution of these data required a staged analytical procedure involving data cleaning, temporal alignment, aggregation, within-participant comparisons, and inferential analysis.

The analytical procedures applied to the facial expression data in the current study were informed by a literature search conducted in the Web of Science database in August 2025 using the search term "FaceReader", which identified 158 English-language journal articles. The bibliographic data of these records provided a useful indication of broader trends in the use of the software: the number of articles explicitly referring to and applying FaceReader increased by 267% between 2015 and 2024.

This survey was supplemented by a case-based review of 14 publications employing FaceReader across different contexts, including education, marketing, social media, trading behaviour, and sport psychology, drawing on the findings of the first review of studies using Noldus FaceReader (Landmann 2023). The studies were purposefully selected to provide insight into the diversity of analytical options and application contexts, including education (e.g. Terzis, Moridis & Economides, 2012; 2013; Chiu, Chou, Wu & Liaw, 2014), social media (Lewinski 2015b), marketing (Neomániová, Berčík and Horská 2018), stock market trading behaviour (Breaban and Noussair 2018) or extreme sport psychology (Hetland et al. 2019) (see Table 5 in the Appendix).

Facial analysis data were analysed in three phases using different strategies:

1. Initial assessment: automatically generated FaceReader graphs
2. Statistical analysis conducted in Excel: analysis of raw values and relative frequency of peak phases
3. Advanced statistical analysis conducted in Python and JASP: within-person analysis of facial analysis and questionnaire data

FaceReader provided a detailed analysis of 73,920 frames from video recordings at 640 × 480 resolution. Numerical results from the questionnaires and automated facial expression detection were analysed using correlation and linear regression, with statistical software such as Python, JASP, and Noldus The Observer XT, which enabled more accurate analysis of the large dataset collected by FaceReader and a more thorough exploration of the relationships between variables. These relationships have been examined using Student paired-samples T-tests or Wilcoxon signed-rank tests, depending on the data distribution.

8.7.3.1 Managing data integrity: cleaning, missing data, and calibration

To reduce bias introduced by individual facial characteristics (Lewinski 2015a), the statistical analysis was preceded by a semi-automated facial expression calibration for each participant's recordings, selecting a neutral frame as a standard reference point for recalculating facial expression results.

Of the 55 participants, six recordings were excluded from facial analysis due to technical issues, including poor camera settings, insufficient data quality from glasses, or inadequate lighting. This resulted in 49 valid facial analyses. One additional participant showed boredom values of 0 throughout; therefore, the participant was excluded only from analyses of the boredom variable (pairwise deletion), while the remaining data were retained.

Outliers were identified using the outlier labelling rule with a multiplier of 2.2 applied to the interquartile range (Hoaglin & Iglewicz, 1987). Observations beyond the resulting cut-off were flagged but retained in the dataset because the non-parametric analyses employed - Wilcoxon signed-rank tests and Spearman's correlations - are robust to deviations from normality and relatively insensitive to extreme values.

Data synchronisation is critical for accurately capturing emotional responses and behavioural patterns in facial analysis (Lewinski 2015b; Terzis, Moridis and Economides 2012). FaceReader recordings are launched a few seconds before the stimulus is displayed. Despite using a FaceReader utility (Stimulus Presentation Tool) to synchronise the instructional video recording and analysis with the stimulus, the length of this initial delay varied by 2-5 seconds across recordings, necessitating a subsequent calculation to

synchronise the exported datasets for each participant, correcting for the delay applied in each recording.

8.7.3.2 Initial assessment by automatically generated FaceReader graphs

The primary purpose of the FaceReader tool is automated facial action analysis, that is, the automated detection and quantification of facial movements and their emotional patterns over time, but it also generates numerical reports, graphs and summaries, supporting the rapid preliminary overview of the data patterns. The data analysis window of FaceReader (see Appendix F) provides line, pie, and bar charts showing the values of the selected variables, enabling a quick comparison of changes in facial expressions with the corresponding stimulus segments. More thorough analyses are typically conducted after exporting the data and processing it further in software such as Excel or statistical packages.

8.7.3.3 Relative frequency analysis of peak boredom moments

Non-normalised values provided by facial analysis should be interpreted with caution, as baseline facial features may affect these scores (Talen & den Uyl, 2022; Hirt et al., 2018). Compared with basic emotions, boredom appears to be more complex, does not appear to be associated with facial expressions that are as universal or as clearly distinguishable as those of basic emotions, and is less straightforward to infer from facial expressions (Gudi 2017; Hirt et al. 2018), especially as its expression may vary with an individual's level of boredom proneness (Melton and Schulenberg 2009; Vodanovich and Kass 2011). Therefore, instead of direct between-subject comparisons, the data were normalised to focus on within-subject changes (Talen & den Uyl, 2022), as recommended by Landmann's overview of FaceReader studies (2023).

Boredom peak phases were identified based on the individual maximum values and average values of facial expression labelled as boredom by FaceReader (Hirt et al. 2018; Lewinski 2015a; Lewinski, Den Uyl and Butler 2014; Talen and Den Uyl 2022): For each participant, the maximum boredom value was identified, and all seconds in which boredom exceeded 90% of this individual maximum were selected. Rather than comparing second-by-second values across participants, the analysis focused on the relative frequency of these individual peak boredom moments. The dataset was then divided into 10-second time windows, given that boredom detection requires a minimum of 5 seconds (Gudi 2017).

The relative frequency of these peak boredom moments within the time windows was examined to identify the most boredom-intensive phases of the video. These phases were then compared with the video content to determine what occurred immediately before and during them.

8.7.3.4 Advanced statistical analysis

8.7.3.5 Paired sample T-tests

The effects of certain communication tools (multimedia principles) on boredom, arousal, valence, and basic emotions were examined using paired-samples t-tests (when data were normally distributed) or the Wilcoxon signed-rank test (when data were not normally distributed). The principles were selected for this analysis based on their prominence and repeated occurrence in the applied stimulus video. Furthermore, all but one of these principles were referenced in at least 30% of the student interviews, further supporting their relevance to the study. Data normality was assessed using the Shapiro-Wilk test.

Confidence intervals (CIs) have been used as error indicators in effect size calculations and bar plots to provide a more informative measure of variability, allowing for direct interpretation of statistical significance and generalizability beyond the sample, unlike standard errors (SEs), which primarily reflect the precision of sample means. When highlighting and reporting on small, medium, or high effect sizes, as defined by Cureton (1956), non-significant results ($p \geq 0.05$) have been excluded from categorisation. Still, they are nevertheless highlighted as an indication of trends.

To summarise the variable-dependent associations between video design tools and learners' affective states, a heatmap has been compiled that contrasts affective indicators across segments in which a given design element was present with those in which it was absent. To reduce construct proliferation, we organised video design features into two theory-informed higher-order categories (Immediacy and Clarity). For each participant and for each design feature (e.g., signalling, embodiment/face, emotional design, redundancy avoidance), FaceReader-derived scores were aggregated separately across (a) all time windows aligned with feature-present segments and (b) all time windows aligned with feature-absent segments. The primary indicators were valence, arousal, boredom, and neutrality; other discrete emotion outputs were treated as secondary descriptors to aid interpretation of patterns. Statistical contrasts were computed as within-participant paired

comparisons between the two segment sets (feature present/Measure 1 vs feature absent/Measure 2), with multiplicity controlled using Holm–Bonferroni adjustment across the family of tests. For each design feature and indicator, we summarised the direction and size of the difference using an ordinal code in the heatmap (e.g., +/++/+++ for positive and −/−−/−−− for negative differences), with the lightest shading indicating contrasts that were not significant after Holm–Bonferroni correction. This representation provides an at-a-glance comparison across cues and indicators when direct numerical comparisons are difficult due to differing score distributions and effect-size metrics, while still preserving the underlying inferential logic of the paired tests and corrected significance decisions.

8.7.3.6 Cross-method comparison of self-report and facial-analysis indicators

The following analyses were conducted to explore the extent to which different multimodal indicators of boredom and interest showed corresponding patterns across self-report and facial-analysis measures.

Comparing FaceReader data with self-reported boredom and interest: As discussed in Section 10.5.2.2, FaceReader's estimates of boredom should be interpreted with caution, as the software's affective-attitude measures are more complex than basic emotion outputs and have shown limited agreement with self-reports in previous research (Gudi, 2017; Lewinski, Fransen et al., 2014). To obtain study-specific evidence on the interpretability of these automated boredom and interest estimates, students were asked, less than 10 minutes after watching the video, to indicate which parts they found most interesting and which they found boring or difficult to follow. Based on these retrospective reports, individual time intervals were identified with split-second precision and labelled as "interesting," "boring," or "difficult to follow." Mean values of boredom, interest, valence, arousal, and neutral facial expression were then compared across the self-reported segments in an exploratory cross-method analysis.

Correlations between questionnaires and facial analysis: The questionnaires completed by students before and after the observation were analysed using correlation analysis to examine relationships among self-reported emotional states, the mean value of automatically detected emotions, and the results of the cognitive and emotional interest scale. Given the low reliability of automated facial analysis in detecting boredom, rather

than the mean values, the mean of the boredom values in the top 10% of the individual boredom range was used.

Examining facial affects before and after the most boring phase: Separate repeated-measures ANOVAs were conducted for FaceReader-derived arousal and valence to examine differences across the four periods surrounding the video section most frequently identified as boring. Participant-level median arousal and valence values were calculated for each period. Sphericity was assessed using Mauchly's test, and Greenhouse–Geisser correction was applied where the assumption of sphericity was violated.

Examining temporal and individual variation in FaceReader-derived boredom: Participant-level median boredom values were calculated for the four-minute episode most frequently identified as boring and for the equally long preceding and following periods. Differences across the three periods were examined using the Friedman test, followed by pairwise Wilcoxon signed-rank tests with Holm correction. Individual trajectories were visualised using a spaghetti plot, and exploratory cluster analysis was used to examine between-participant variation in boredom profiles.

9 RESULTS

The current chapter reports the main patterns observed in the collected data and summarises the analysis results, in line with the analytic expectations formulated in Section 1.3. This overview is introduced in Section 9.1, which presents the qualitative content analysis of the instructional video.

Throughout this chapter, references to “boredom”, “happy”, “sad”, “angry”, and other emotion categories derived from FaceReader denote software-generated facial-analysis indicators unless explicitly identified as self-reported experiences. Statistical differences in these indicators are interpreted as associations between instructional conditions and classified facial patterns; they do not, by themselves, establish changes in participants' subjectively experienced emotions.

9.1 Application of the coding framework to the instructional video

An important step in preparing for the statistical analysis was coding the instructional video to identify the temporal occurrence of each multimedia design feature with one-second precision. A detailed description of the video, including a slide-by-slide overview of the communication techniques, topics and visual elements employed, is provided in Table 10 in Section 8.3. The coding framework applied in the stimulus analysis is presented in Table 12 in Section 8.7.1.

The video was manually coded to mark the intervals in which specific design principles were present. Each feature was coded for every second using a binary scheme (presence = 1; absence = 0). The coding process was supported by Noldus TheObserver XT, enabling frame-by-frame assessment with sub-second precision. Figure 12 shows the temporal distribution of coded design features across the video. Based on this coding, feature-specific datasets were generated in numerical form and subsequently analysed in Python, Excel, and JASP (see Section 9.2.3). The coding used mutually exclusive categories, resulting in binary coding for each second, except for segmentation, for which only the first half-minute of each slide was coded as segmented. The figure demonstrates that, with a few exceptions, the phases characterised by a given multimedia principle or communication tool were typically very short, underscoring the necessity of fine-grained numerical analysis.

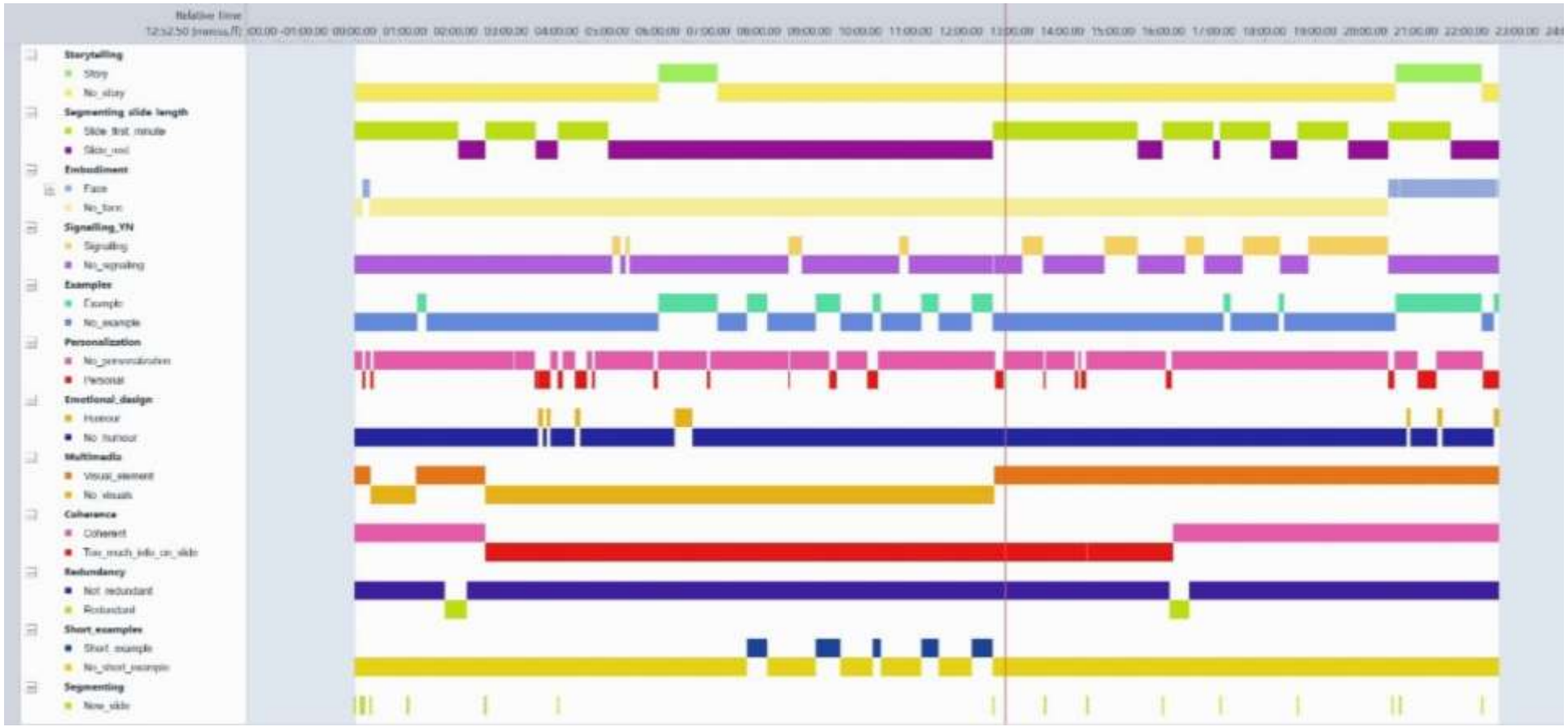


Figure 12: Duration of certain multimedia principles applied in the instructional video.

(Author's own compilation using Noldus TheObserver XT 15)

9.2 RQ1: Design features most strongly associated with boredom

This section presents the findings most directly relevant to Analytic Expectation 1, which hypothesised that multimedia design principles are associated with a lower level of boredom, and that segmenting, signalling, personalisation, humour and the appearance of the instructor's face show a particularly strong negative association with boredom.

Analysis of facial expression data followed a multi-stage, multimethod strategy, reflecting the iterative development of the analytical procedure throughout the research. The three sub-chapters, therefore, represent three different stages and levels of data analysis. Although the strategy presented in Section 10.2.3 provides the most accurate and reliable results, Sections 10.2.1 and 10.2.2 offer insight into how the final analytical method was developed. They also illustrate the relative precision of the earlier analytical attempts and the extent to which the results obtained at different stages were consistent with one another.

The three stages differed not only in their analytical procedures but also in the temporal granularity of their analyses. The first stage relied on 30-second time windows, the second on 10-second segments, and the third on a 1-second resolution. The first stage provided an initial overview based on the early recordings available at that point, while the second stage served as an intermediate step in refining the analytical logic. The third stage, applied to the full dataset, enabled a more precise and statistically more robust examination of the observed patterns. The purpose of reporting all these stages was to document the methodological refinement process that led to the final analytical solution: the earlier stages primarily played a developmental role, whereas the final stage provided the basis for the main statistical analyses reported in detail.

9.2.1 Raw mean values of boredom and interest

At an early stage of the data collection process, as a preliminary attempt to identify patterns of boredom and interest, changes in the aggregated, non-normalised raw mean values of boredom and interest (as measured by FaceReader) were examined (N=25). This analysis aimed to detect periods of elevated boredom and elevated interest. As Figure 13 shows, changes of mean boredom and interest values were not consistently in opposite directions; that is, low interest was not always associated with high boredom,

and vice versa.

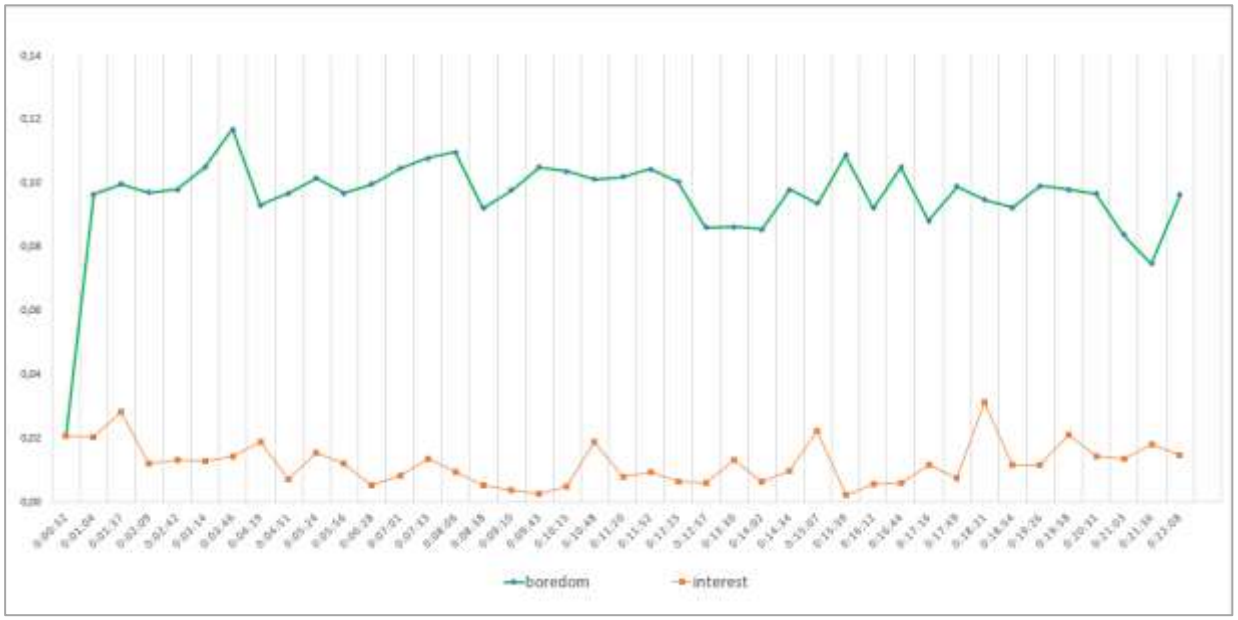


Figure 13: Changes in the mean raw values of boredom and interest while viewing a 22.6-minute video presentation

Note. Measured by facial expression analysis (N = 25) and calculated over 30-second intervals. The lower number of participants in this analysis reflects the fact that it was conducted at an earlier stage of the research.

It was nevertheless hypothesised that segments characterised by the simultaneous occurrence of high boredom and low interest indicated that the respective segment of the instructional video triggered boredom in the majority of participants, whereas segments characterised by low mean boredom and high mean interest were associated with a low level of boredom. Based on these criteria, facial analysis data on boredom and interest collected from students during the 22.6-minute video were examined.

To assess the synchronous and asynchronous fluctuations of boredom and interest, the aim was to identify congruent patterns: *elevated boredom* segments (simultaneously high boredom and low interest) and *elevated interest* segments (low cumulative mean boredom and high interest). To ensure a clearer interpretation of these segments, only windows showing both congruence and relatively low variability were included in this part of the analysis. Variability was assessed separately for each 30-second window by calculating the standard deviation for that window (SD_{window}), and comparing it with the mean standard

deviation across all windows (M_{SD}) (D'Mello, Dieterle and Duckworth 2017; Gudi 2017; Hirt et al. 2018; Landmann 2023). Windows were classified as low-variance when

$$SD_{window} < M_{SD}$$

and as high-variance otherwise. The overall mean boredom value across windows was 0.0973, and the standard deviation of the window means was 0.0082; therefore, the threshold for elevated boredom was defined as

$$T_{boredom} = M_{boredom} + 0.4 \times SD_{boredom} = 0.1006$$

For interest, the overall mean was 0.0120, and the standard deviation across window means was 0.0067, resulting in the following threshold:

$$T_{interest} = M_{interest} + 0.4 \times SD_{interest} = 0.0147$$

These thresholds were used to distinguish elevated from non-elevated boredom and interest, while the variance filter was applied separately to exclude unstable windows.

Although phases identified with this method could be determined only with 30-second accuracy, this level of precision was sufficient to identify 5 elevated boredom windows, whereas no elevated interest windows were identified because no time window simultaneously exceeded the interest threshold and met the low-variability criterion.

Table 13 presents the video features associated with elevated boredom phases, listed in order of mean boredom intensity. The results suggest that most elevated boredom phases occurred during Segment 6, which involved explaining a table displayed for more than 8 minutes.

Table 13: Video features associated with congruent patterns of viewer boredom and interest in the 30-second time-window analysis (N = 25)

$Time_{window}$ (mm:ss)	**$Mean_{boredom}$**	**SD_{window}**	**Segment number**[5]	**Segment features**	

[5] As defined in Table 10 (Section 8.3)

07:33–08:06	0.1096	0.0565	6	Table-structured English text, with on-screen highlighting during Hungarian narration	
02:42–03:14	0.1050	0.0569	5		
06:28–07:01	0.1046	0.0527	6	Table-structured English text, with on-screen highlighting during Hungarian narration	
09:43–10:15	0.1037	0.0582	6		
04:51–05:24	0.1015	0.0559	6		

Note. Ordered by mean boredom intensity

The results were consistent with the interview findings; however, they did not provide sufficient information to estimate the most likely triggers of boredom during this phase. At the same time, the substantial individual variability in boredom and interest values indicated the necessity of an analytical approach that incorporated or was preceded by data standardisation.

9.2.2 Relative frequency analysis of boredom peak moments

The next step in analysing the numerical data exported from the FaceReader analysis was a relative frequency analysis, as discussed in Section 8.7.3.3. This analysis improved the comparability of the results by examining, across the full sample and the entire video, how often 10-second time windows exceeded the participant's individual top 10% boredom threshold. Figure 14 presents the number of participants showing "boredom peak" moments across 10-second time windows. Blue bars indicate the relative frequency of peak boredom moments, red arrows mark the windows identified as high-boredom phases, and green bands indicate periods without boredom peaks used as comparison segments.

Episodes preceding the moments when the highest number of students showed expressions of boredom exceeding 90% of their individual maximum included the following segments (a detailed description of the video segments is provided in Table 10 in Section 8.3):

- slide incoherent with the spoken narration (Segment 6);
- slide with too much detailed and extraneous information (Segment 8);
- text without any illustration (Segment 3);
- slides displayed for more than 5 minutes without an illustration or a human face (Segment 6).

The features that preceded and co-occurred with segments in which none of the participants showed boredom peak values (indicated by green bands in Figure 14) included 5 phases characterised by the features discussed in Table 14.

Order of occurrence	Time window-length	Features	Segments, as categorised in Table 10
1	130 seconds	signalling, personalisation, humour, segmenting, storytelling, example	end of Segment 5, initial part of Segment 6
2	60 seconds	visual elements, signalling, segmenting, coherence (avoiding extraneous information)	Segment 7, initial part of Segment 8
3	30 seconds	visual elements, signalling, coherence (avoiding extraneous information)	Segment 12
4	20 seconds	presenter's face, display of a singer's smiling portrait, storytelling	Segment 13-14
5	30 seconds	presenter's face, humour, storytelling, smile	Segment 15

Table 14: Features of phases without boredom peak moments identified through relative frequency analysis

As the video progresses, these low-boredom phases generally become shorter, despite the increased frequency of attention-maintaining features in the second phase, such as more frequent slide changes, graphs, examples and storytelling.

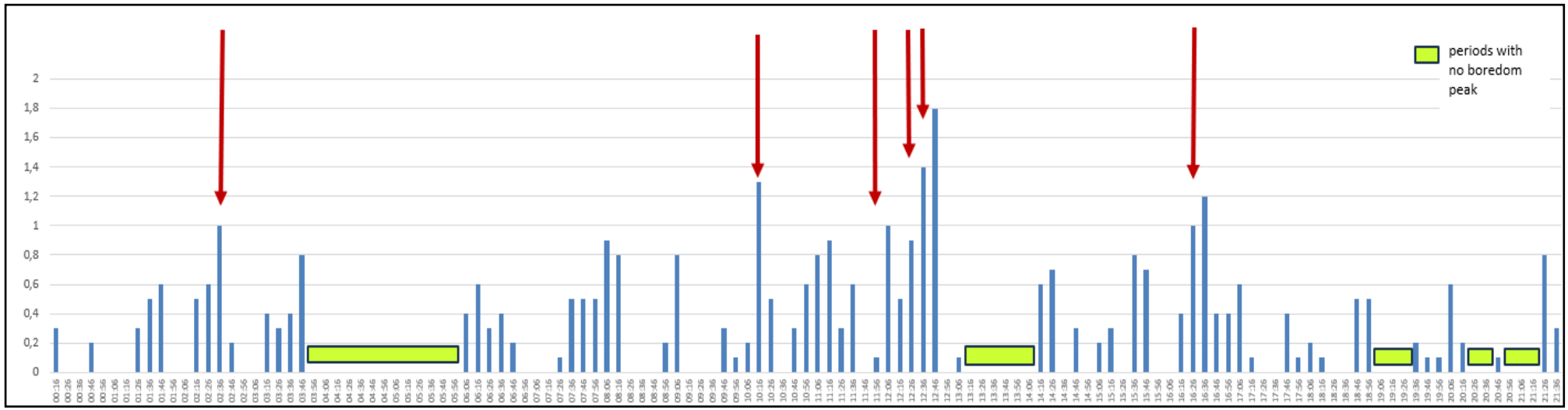


Figure 14: Boredom peak phases per 10-second time windows (N=49)

9.2.3 Associations of multimedia learning features with boredom and neutral indicators

Noldus FaceReader includes a feature (custom expression “boredom”) that provides estimated values for boredom (see Section 7.5), but its sensitivity is relatively low at 55%. As demonstrated in the relevant literature, boredom is most commonly characterised by low arousal and low valence (Goetz et al. 2024). When assessing levels of boredom, this section also considers the values for arousal, valence, and neutral facial expressions as indicative reference values. Additionally, based on the findings of FaceReader's custom expressions validation report (Gudi 2017) and of the FaceReader methodology review (Landmann 2023), it was assumed that neutral values may partly reflect missed boredom, as boredom may appear in facial patterns that are difficult to distinguish from neutral states. Therefore, the current section discusses values for *neutral* alongside those for boredom, suggesting that some cases of boredom may not have been detected.

As mentioned in Section 8.7.3.5, the paired samples test was selected based on the distribution of the results. For normally distributed data ($p > 0.05$), Student’s t-test was used, while the Wilcoxon Signed Rank test was used when the distribution of the result values deviated from normality ($p < 0.1$). In cases where the p-value fell within the range $0.05 < p < 0.10.05 < p < 0.10.05 < p < 0.1$, the data distribution was assessed as tending towards normality. Normality was tested with the Shapiro-Wilk test (see Table 15). The distributions of boredom values were normal for all variables except embodiment (face – no face) and redundancy, which deviated from normality, and smile, which only tended towards normality. Accordingly, these variables were analysed using the Wilcoxon signed-rank test.

Table 15: Assumption test for normality for the calculation of the T-test

Test of Normality (Shapiro-Wilk)

			W	p
Boredom | No face	-	Boredom | Face	0.952	0.043
Boredom | No signalling	-	Boredom | Signalling	0.987	0.847
Boredom | No personalisation	-	Boredom | Personal	0.974	0.352
Boredom | No visuals	-	Boredom | Visual element	0.971	0.265
Boredom | No story	-	Boredom | Story	0.988	0.885
Boredom | Not Coherent	-	Boredom | Coherent	0.967	0.182
Boredom | Redundant	-	Boredom | Not redundant	0.842	< .001
Boredom | Slide rest	-	Boredom | Slide first minute	0.975	0.364
Boredom | No humour	-	Boredom | Humour	0.970	0.245
Boredom | No example	-	Boredom | Example	0.979	0.544
Boredom | No Smile	-	Boredom | Smile	0.957	0.073

The alternative hypothesis formulated in JASP stated that boredom would be higher during segments in which the given multimedia tool was absent (Measure 1) than during segments in which it was present (Measure 2):

$$H_a: Mean_{Measure\ 1} > Mean_{Measure\ 2}$$

In the case of facial expression of "boredom", the effect sizes were strongest for "face" (r_{rb} = 0.647) visual elements (d = 0.794), coherence (d = 0.758), and medium effect size was shown in the case of segmenting (slide first minute) (d = 0.540), smile (r_{rb} = 0.329) and personalization (d = 0.503)[6] (see Figure 15 and

Table 16).

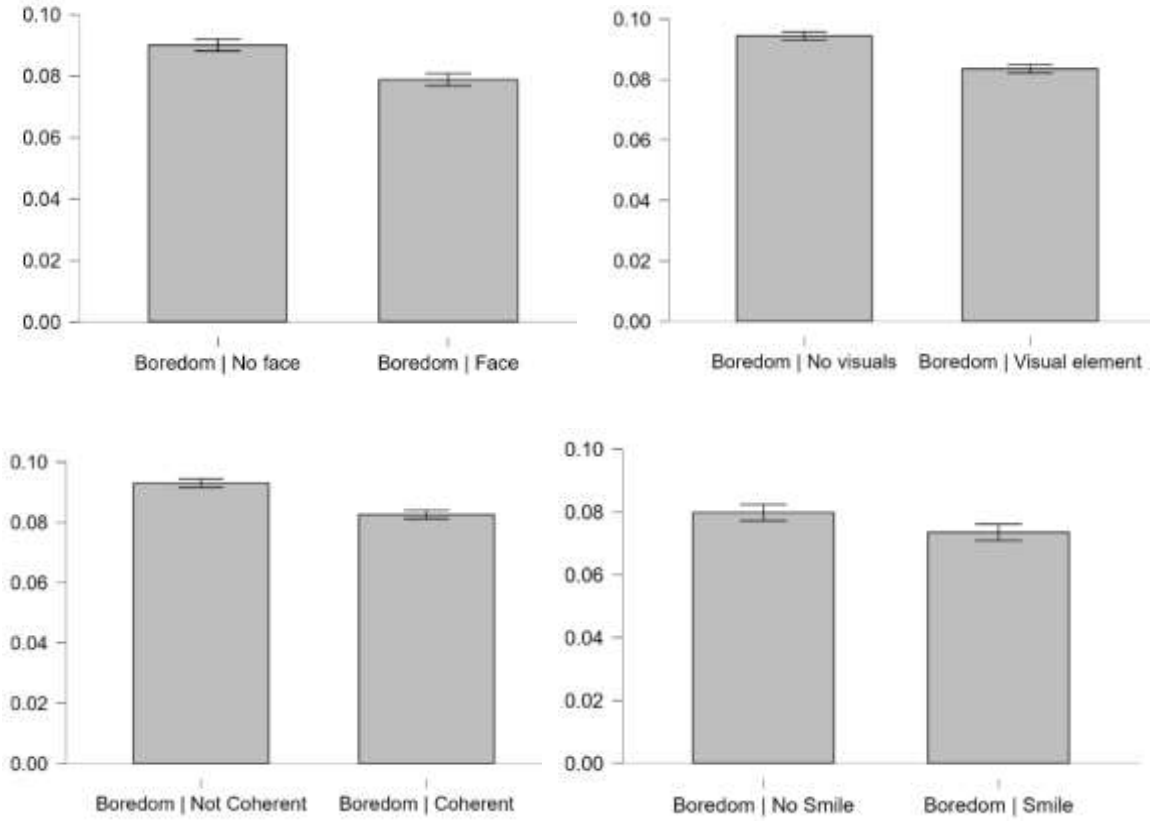


Figure 15: Associations between selected multimedia-learning features and FaceReader-labelled boredom

[6] Student's t-test was used for normally distributed variables (p > .05), whereas the Wilcoxon signed-rank test was used when normality was violated or only marginal (0.05 < p < 0.10). Normality was tested with the Shapiro–Wilk test

Error bars represent 95% confidence intervals for the means. JASP Team (2024). JASP (Version 0.19.2) [Computer software].

Table 16: Associations between selected multimedia-learning features and FaceReader-labelled boredom (paired-samples t-tests and Wilcoxon signed-rank tests)

Measure 1		Measure 2	Test	Statistic	z	df	p	rank	Adjusted α	Adjusted Significance	Holm-Bonferroni corrected p	Effect Size	SE Effect Size	95% CI for Effect Size	
														Lower	Upper
Boredom | No visuals	-	Boredom | Visual element	Student	5.56		48	0.000001	1	0.0045	Significant	0.000006	0.794	0.034	0.521	∞
Boredom | Not Coherent	-	Boredom | Coherent	Student	5.31		48	0.000001	2	0.0050	Significant	0.000014	0.758	0.033	0.487	∞
Boredom | No face		Boredom | Face	Wilcoxon	1009.00	3.944		0.000019	3	0.0056	Significant	0.000168	0.647	0.163	0.463	∞
Boredom | Slide rest	-	Boredom | Slide first minute	Student	3.78		48	0.000218	4	0.0063	Significant	0.001742	0.540	0.026	0.285	∞
Boredom | No personalisation	-	Boredom | Personal	Student	3.52		48	0.000474	5	0.0071	Significant	0.003318	0.503	0.023	0.251	∞
Boredom | No signalling	-	Boredom | Signalling	Student	2.70		48	0.005000	6	0.0083	Significant	0.030000	0.385	0.032	0.14	∞
Boredom | No Smile		Boredom | Smile	Wilcoxon	688.00	1.925		0.027000	7	0.0100	Not Significant	0.135000	0.329	0.17	0.061	∞
Boredom | No humour	-	Boredom | Humour	Student	1.80		48	0.039000	8	0.0125	Not Significant	0.156000	0.257	0.05	0.017	∞
Boredom | No story	-	Boredom | Story	Student	1.19		48	0.120000	9	0.0167	Not Significant	0.360000	0.170	0.041	-0.068	∞
Boredom | Redundant		Boredom | Not redundant	Wilcoxon	605.00	-0.075		0.531000	10	0.0250	Not Significant	1.000000	-0.012	0.163	-0.275	∞
Boredom | No example	-	Boredom | Example	Student	-1.08		48	0.858000	11	0.0500	Not Significant	0.858000	-0.155	0.028	-0.39	∞

Note. For all tests, the alternative hypothesis specifies that Measure 1 is greater than Measure 2. For example, Boredom | No signalling is greater than Boredom | Signalling.

Note. For the Student t-test, the effect size is given by Cohen's d . For the Wilcoxon test, the effect size is given by the matched rank biserial correlation.

Note. Shaded cells with grey indicate small (r_{rb} > 0.1), medium (r_{rb} > 0.3) and large (r_{rb} > 0.5) effect sizes, as defined by (Cureton, 1956) for the Wilcoxon Signed-Rank Test. Cohen's d values indicate small (d > 0.2), medium (d > 0.5), and large (d > 0.8) effect sizes, as defined by Cohen (1988). The lightest grey indicates non-significant but noteworthy results (p ≥ 0.05).

According to Table 1 of Appendix G, features showing the largest between-condition differences/effect sizes for "neutral" included visual elements (r_{rb} = 0.671), lack of redundancy (r_{rb} = 0.553) and signalling (r_{rb} = 0.485).

For *valence* and *arousal*, most variables were not normally distributed. Therefore, the Student's t-test could be applied to only four variable pairs, while the Wilcoxon signed-rank test was used for the remaining comparisons.

Table 2 and Table 3 in Appendix D display the outcomes of the two types of T-tests, sorted by effect size. For information supporting the interpretation of the effect sizes, see the footnotes of the table.

The alternative hypothesis, based on literature findings and student interviews, was that the mean arousal and valence calculated for Measure 2 (i.e., during segments where the given multimedia principle was applied) differ from those of the first variable.

$$H_a: \text{Mean}_{\text{Measure 1}} < \text{Mean}_{\text{Measure 2}}$$

Effect sizes were therefore calculated by subtracting the tool-use condition (Measure 2) values from the baseline (Measure 1) values. Negative effect sizes for valence indicate higher valence values during the tool-use condition, while negative effect-size values for arousal similarly indicate higher arousal values during the tool-use condition.

Valence showed large effect sizes for humour (r_{rb} = -0.82), smile (r_{rb} = -0.78), and the appearance of the lecturer's face (r_{rb} = -0.573) and personalisation (d = -0.870). Similarly, for *arousal*, large effect sizes were observed for personalisation (d = -0.843), humour (r_{rb} = -0.680), appearance of the instructor's face (r_{rb} = -0.645), and slide's first minute (r_{rb} = -0.504)[7] (see Tables 8 and 9 in Appendix F).

Results discussed in the current section indicate that statistically significant within-participant differences were observed for several multimedia learning principles in relation

[7] The negative effect sizes indicate that the values measured during the use of the given tool, were higher compared to those measured during the baseline (or tool-free) condition, as the effect size is calculated by subtracting the principle-use condition values from the baseline values.

to boredom and to cognitive-affective states associated with higher arousal and valence; however, the strongest effect sizes were consistently observed for personalisation, humour, and the instructor's face (embodiment), with segmenting (slide first minute) also showing a particularly large effect size for arousal.

9.2.4 Students' self-perceived boredom - Interviews

9.2.4.1 Students' reports on sections perceived as "interesting" and "boring"

The frequency of multimedia design elements mentioned in student retrospective interviews as supporting or impairing their attention is shown in Figure 16 and Figure 17.

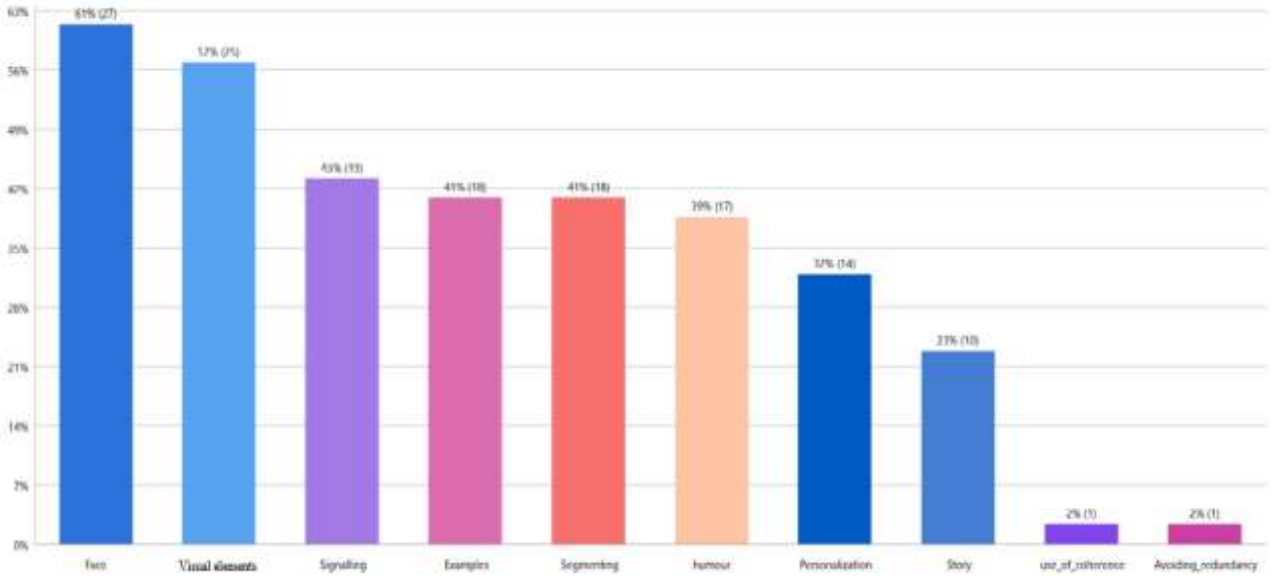


Figure 16: Attention-supporting communication features mentioned by participants

Note. The chart shows the proportion of participants whose interviews contained references to each coded communication feature. Percentages represent the proportion of participants who mentioned the feature at least once during the interview (N = 44). Codes were generated through qualitative content analysis using MAXQDA.

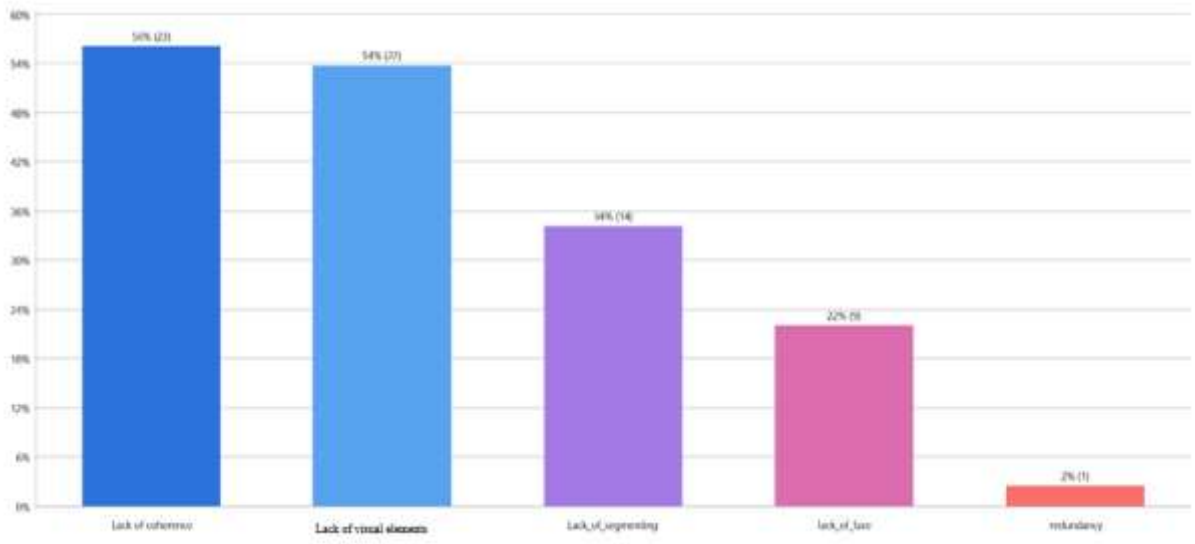


Figure 17: Attention-impairing communication features mentioned by participants

Note. The chart shows the proportion of participants whose interviews contained references to each coded communication feature. Percentages represent the proportion of participants who reported that the feature impaired their attention and learning at least once during the interview (N = 44). Codes were generated through qualitative content analysis using MAXQDA.

When comparing the above features across their higher-order categories, i. e., immediacy- and clarity-related strategies (see Section 5.2), the proportions of the mentioned features were relatively balanced, as shown in Table 17.

Table 17

Table 17: Number and proportion of participants mentioning immediacy and clarity-related features perceived as supporting their attention.

	Frequency[8]	**Percentage**
Clarity-related features	40	90.91
Immediacy-related features	38	86.36
ANALYSED DOCUMENTS	44	100.00

[8] Number of students referring to certain video design features (N=44)

Taken together, the interview findings illustrated both convergent and divergent patterns across automated facial indicators and students' self-reported experiences.

9.2.4.2 Boredom and recall difficulty

The interviews also revealed a pattern beyond the scope of the research questions: several students reported that the sections they found boring were also more difficult to recall, as reflected in the excerpts below. Some responses to the question, *What was the most difficult to follow?*, were as follows:

"*Maybe at the very beginning, where... I can't even recall... what it was about exactly*"

"*Uh...*, *the one I can't remember what you were talking about..., wait...* [laughter]... *is that one of those what kind of graph? Not a column, but a bar chart*…?"

Several of them attributed their recall difficulties to factors unrelated to the course material:

"*…a few times my attention has wandered a bit, but that's partly because I'm tired*"

"*I've been trying really hard to understand what it's about, but I just didn't get it. Maybe it's because of everything that's happened to me today*."

"*I know it's a bit of a morning for me, but that's just my own stupidity.*"

Not only did this phenomenon illustrate the low awareness of boredom as an experience, but it also supported the positive association between boredom and attention and recall problems, as well as cognitive errors, reported in previous research. This pattern can occur for two potential reasons, depending on student competencies and situation: (a) boredom can emerge in response to *cognitively challenging information* (Pekrun et al. 2010; Sweller 1988; Wallace, Vodanovich and Restino 2003), but (b) boredom itself may also *impair* attention and, consequently, memory performance (Goetz et al. 2023, 2024; Pekrun, Marsh, Elliot, et al. 2023). In practice, these processes likely interact reciprocally: cognitive difficulty may trigger boredom, which further impairs attention and memory, making re-engagement increasingly difficult (Hunter and Eastwood 2021; Smith and Tinto 2022).

9.2.4.3 Facial affects before, during and following a high-boredom phase

The video section deemed most boring by most of the interviewees was the closing part of Segment 6 (see Table 10), an 8-minute explanation of the main concepts, during which

only a table of key terms was displayed, accompanied by spoken narration. Although the teacher used strategies to maintain attention during this phase, such as storytelling, humour, examples, and signalling, most students reported losing focus. By selecting and analysing facial data from the last minute of this section, it was possible to examine group-level changes in arousal and valence before, during, and after a period of high boredom.

The dataset of students' facial data has been divided into 4 parts, as shown in Figure 18.

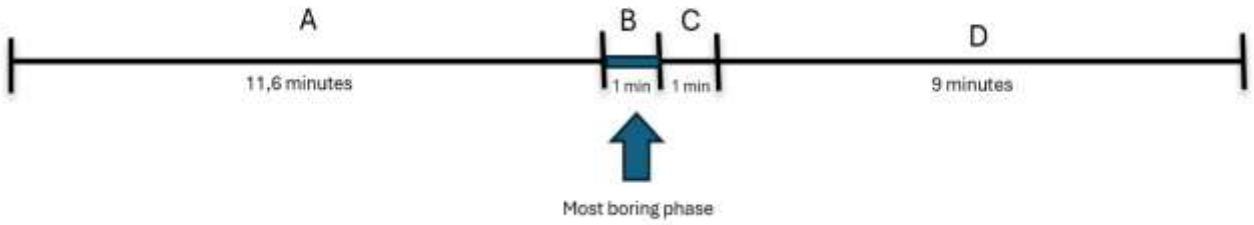


Figure 18: Segmentation of the video according to boredom phases

The one-second facial affect values measured by FaceReader were examined across the periods surrounding the section that participants most frequently identified as boring. As shown in Figure 18, arousal was lower during period B than in the preceding period and returned to higher values in period C. Valence also showed lower values during period B and remained below its initial level in the subsequent periods. Thus, within the present context, the section most frequently identified as boring co-occurred with changes in FaceReader-derived arousal and valence, although the two indicators did not follow identical temporal patterns.

A repeated-measures ANOVA showed a significant overall difference in arousal across the four periods. As Mauchly's test indicated a violation of sphericity, Greenhouse–Geisser correction was applied, F (2.32, 113.63) = 4.24, p = .013. For valence, Mauchly's test did not indicate a violation of sphericity, and the overall difference across the four periods did not reach statistical significance, F (3, 147) = 2.33, p = .077. Thus, the results provide evidence of temporal variation in arousal around the period most frequently identified as boring, while the corresponding variation in valence remained descriptive rather than statistically significant.

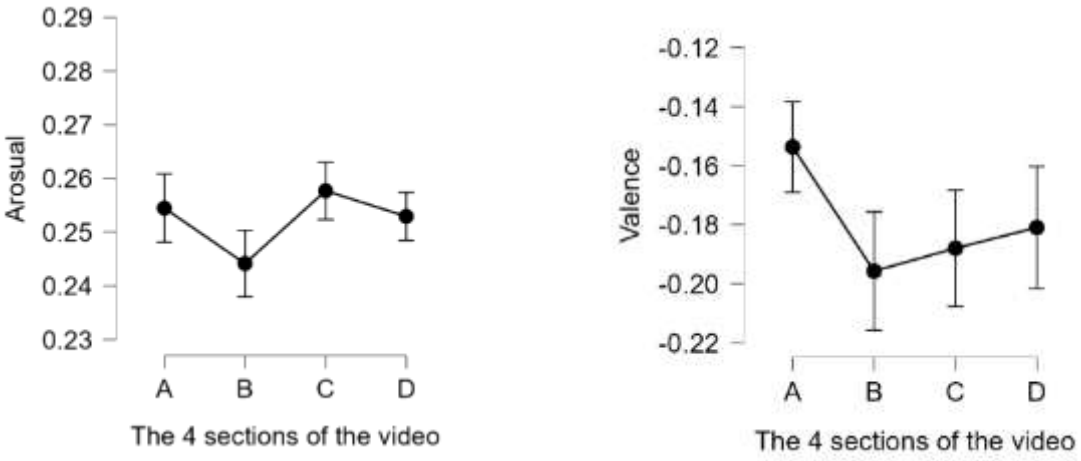


Figure 19: Changes in the mean level of arousal and valence before and after the video segment deemed most boring by students

Another noteworthy feature of the above analysis is that following the decrease of the aggregated mean valence during phase B, valence *did not return to its previous level*, even after the factor causing the boredom changed, namely, during the C and D phases.

To examine whether FaceReader-derived boredom changed around the video segment most frequently identified by participants as boring, participant-level median boredom values were compared across three equal-length periods: the four-minute boring segment, the preceding four minutes, and the following four minutes. Median boredom was .094 before the segment, .094 during the boring segment, and .086 afterwards. Despite the similar medians, the distributions differed significantly across the three periods, Friedman $\chi^2(2) = 16.21$, $p < .001$. Pairwise Wilcoxon signed-rank tests indicated a significant difference between the pre-boredom and boredom periods ($p = .030$) and between the boredom and post-boredom periods ($p < .001$), whereas the pre-boredom and post-boredom periods did not differ significantly ($p = .321$). As illustrated in Figure X, however, individual trajectories showed substantial variability.

9.2.4.4 Individual differences and temporal changes in FaceReader-derived boredom

To examine FaceReader-derived boredom around the episode most frequently identified as boring, participant-level median boredom values were calculated for three equal four-minute periods: the boring episode itself and the immediately preceding and following

periods. A Friedman test indicated a significant overall difference across the three periods, $\chi^2(2) = 16.21$, $p < .001$. Holm-adjusted pairwise Wilcoxon signed-rank tests showed a significant difference between the boredom and post-boredom periods ($p < .001$). The pre-boredom–boredom difference did not reach statistical significance after correction ($p = .060$), and the pre-boredom and post-boredom periods did not differ significantly ($p = .321$). Individual trajectories are presented in Figure 20

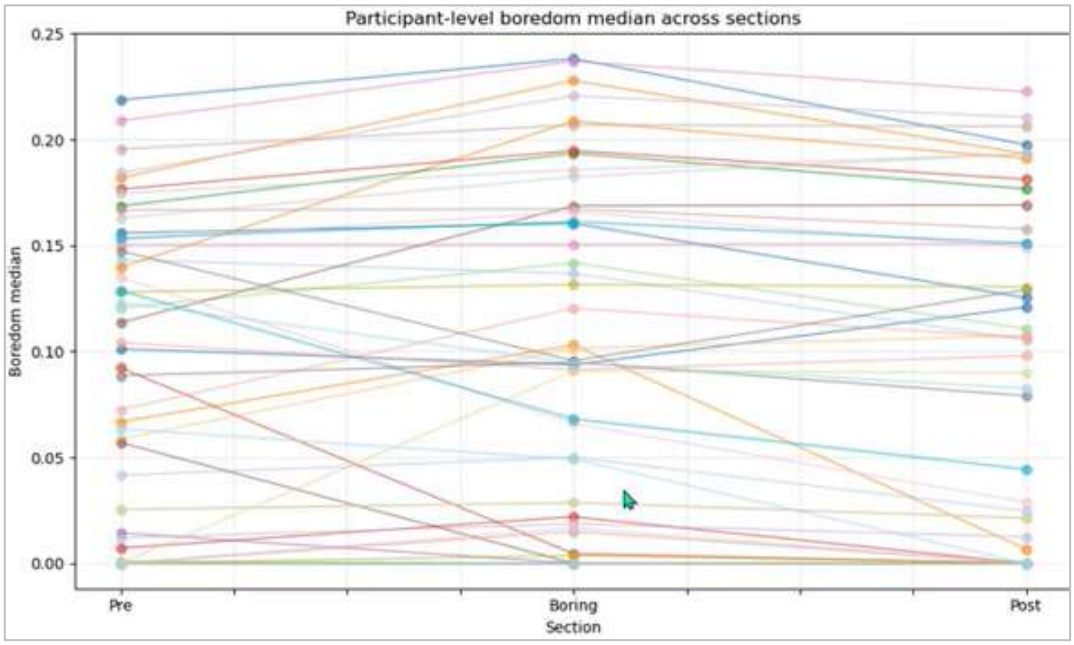


Figure 20: Individual trajectories of FaceReader-derived boredom before, during, and after the video segment most frequently identified as boring

9.3 RQ2: Co-occurrence between design features and affective response patterns

In line with Analytical Expectation 2, this section provides an overview of the relationships between the intensity of expressions identified as specific emotions and the presence of multimedia features. These relationships are based on the within-subject comparison of facial expression intensity values during seconds when selected video design features were present or absent. While Section 9.3.1 provides a brief overview of the automated analysis generated by FaceReader and highlights only a few particularly strong associations, Section 9.3.2 presents a comprehensive heatmap of the relationships between video design features and affective states.

9.3.1 Insight into some selected relationships between design features and basic emotions

Initial assessment by automatically generated FaceReader graphs

As indicated in Section 8.7.3.2, FaceReader's data analysis windows provide immediate access to various customisable graphs, charts and tables that enable a quick overview of the results and reveal major patterns. The aggregated mean values shown on this interface indicated that "neutral", "sad", and "angry" were the most dominant facial expressions (see Figure 21 and Figure 22).

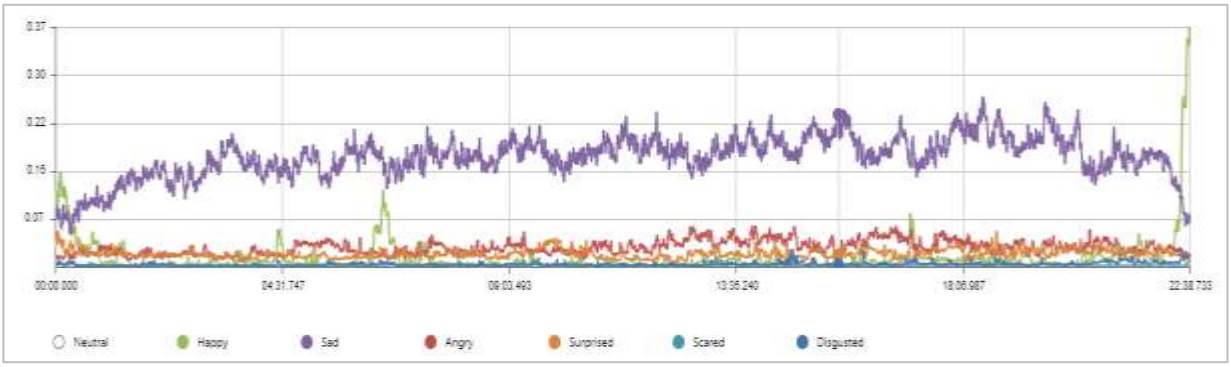


Figure 21: Aggregated mean values of students' emotions on facial expressions (N=50). Noldus FaceReader 9.1 data.

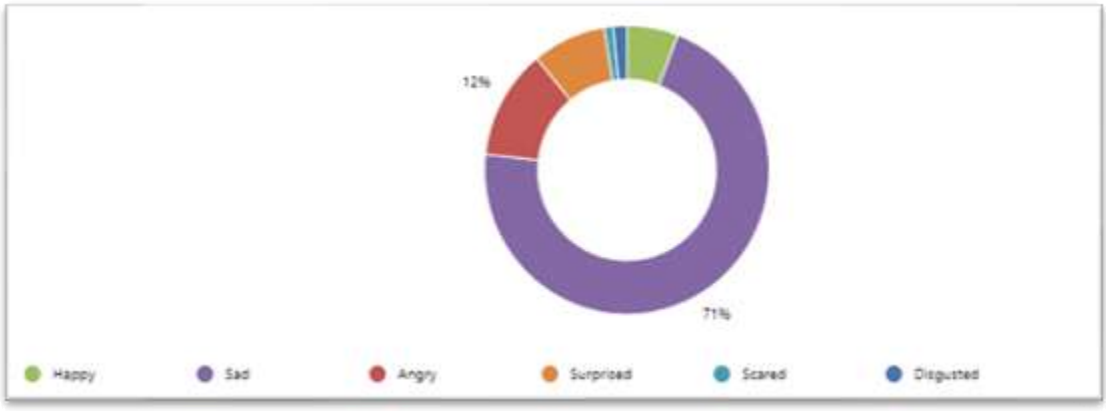


Figure 22: Distribution of dominant expressions, excluding "neutral" (N=50). Noldus FaceReader 9.1 data.

Effect of video design principles on basic emotions

Effect sizes of the selected multimedia measures in relation to the basic emotions showed the largest values for “happy”, “angry” and “sad”. Values indicating “happy” differed significantly from those of other basic emotions because it was the only positive emotion and showed the largest effects across the highest number of multimedia features (see Table 1 of Appendix G). It also stood out in the strength and direction of these effects. The largest effects were found for “face” (r_{rb} = -0.868), “humour” (r_{rb} = -0.834), “personalisation” (r_{rb} = -0.812), and the “absence of redundancy” (r_{rb} = -0.812).[9] These results aligned with the interview outcomes regarding the most influential feature, namely “face”, and factors assessed as attractive by most interviewees, i.e., visual elements (r_{rb} = -0.628), example (r_{rb} = -0.556), coherence (r_{rb} = -0.724) and humour (r_{rb} = -0.834), were among the variables that showed significant and large effect on “happy”.

One of the variables, the design feature "signalling", showed a markedly different pattern compared to the other multimedia principles: it was the only one showing a large negative effect size (r_{rb} = 0.722), strikingly diverging from the others. Further insights into this pattern are provided in Section 9.4.2.1.

9.3.2 Overview of main patterns

Because effect sizes from paired-samples t-tests and Wilcoxon signed-rank tests cannot be directly compared across variables due to differing distributions and metrics, a heatmap (Figure 23) provides an overview of how individual multimedia features relate to affective patterns. This visualisation supports clearer interpretation while reducing the risk of misleading comparisons (Bradley et al. 2012). Representing the relationships calculated with the methods of paired-samples t-tests and Wilcoxon signed-rank tests, as explained in Section 8.7.3.4 provides a detailed overview of the effects of the different features, organised into the higher-level categories of immediacy and clarity. Detailed data behind the heatmap are discussed in Sections 9.1, 10.2 and 10. 3.

[9] In the case of the basic emotions, most datasets did not follow a normal distribution; however, normality was retained for some variables. Therefore, effect sizes were calculated using two-tailed Wilcoxon signed-rank tests for non-normal datasets and paired-samples Student t-tests for normally distributed ones ($p<0.001$).

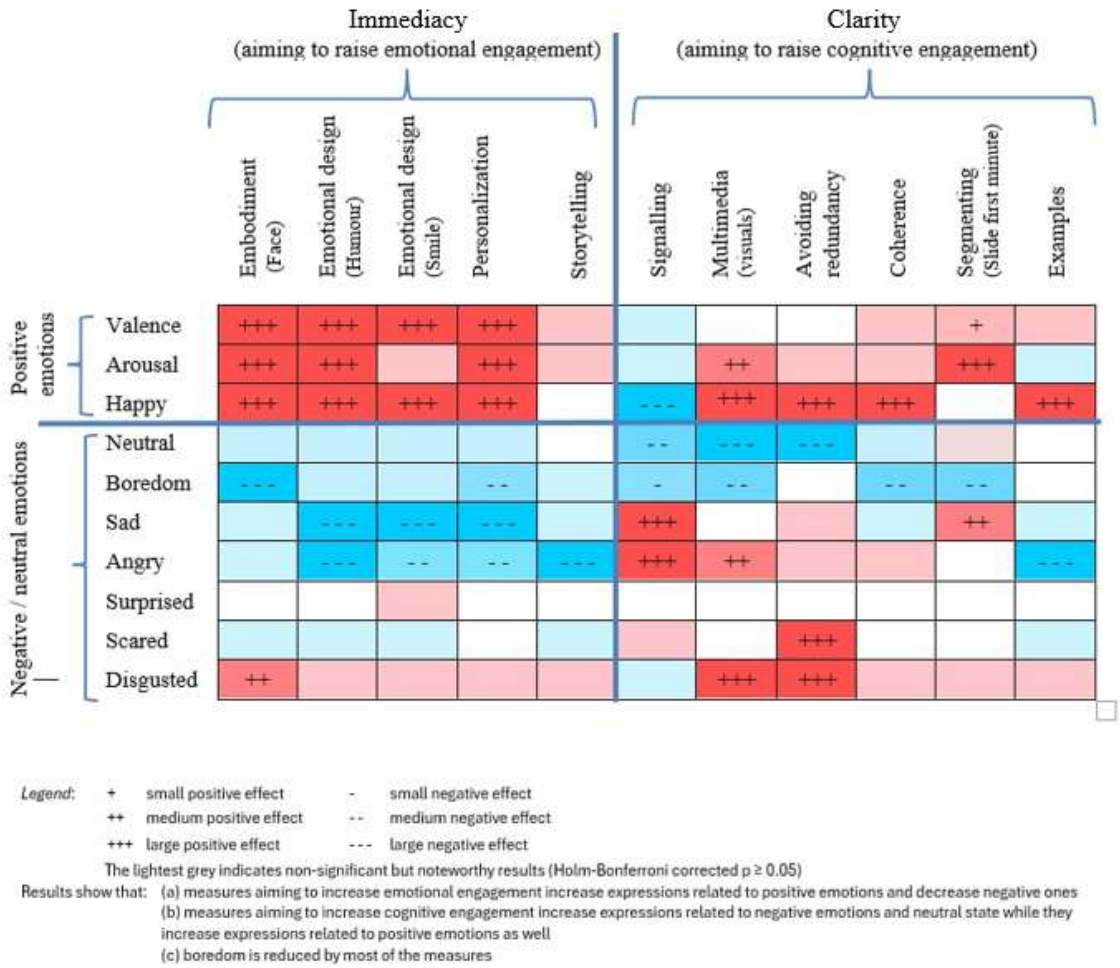


Figure 23: Summary heatmap of the relationship between affective states and the application of certain multimedia principles

- One of the tendencies most clearly demonstrated by the heatmap is the negative correlation between most video design principles and the facial expression of "boredom" and "neutral" - a relationship discussed in Section 9.2.3.
- On the other hand, most features aiming to support *cognitive engagement*, also referred to as *clarity-type interventions* (see Section 5.2.2), similarly, co-occurred with a significant increase of most facial expressions related to affective states promoting learning effectiveness, such as "happy", arousal and valence (Pekrun, Marsh, Elliot, et al. 2023).
- A third tendency, contrary to initial expectations, showed that clarity-related features also co-occurred with higher intensity of facial expressions identified as negative emotions, primarily "angry", "sad", "scared", and "disgusted".

The co-occurrence of facial expressions labelled by FaceReader as “surprised” and “disgusted” showed different patterns from the above-mentioned ones: while most features co-occurred with a higher intensity of “disgusted“, none of the features co-occurred with a large intensity of “surprise”.

Taken together, the results related to RQ2 are broadly consistent with AE2: effect sizes indicated that higher arousal and valence values were significantly associated with the presence of several multimedia design features, with larger effect sizes generally observed for immediacy-related than for clarity-related features. A significant and unexpected finding was the positive association between clarity-related features and facial expressions labelled as negative emotions.

9.4 Alignment across data sources

This section reports the findings relevant to Analytic Expectation 3, focusing on convergence and divergence between the automated facial-expression indicators and questionnaire- and interview-based accounts of boredom and interest. The discrepancies observed across measurement channels are theoretically informative in light of the epistemic position outlined in Section 1.5.1. Self-reports, retrospective interviews, and facial classifications do not provide interchangeable representations of a single directly observable emotional state. They differ in temporal resolution, degree of conscious access, mode of expression, and inferential distance from experience. Convergence, therefore, provides one type of evidence, while divergence can reveal limitations in a particular operationalisation or differences between experienced, remembered, expressed, and automatically classified responses (Barrett et al. 2019; D’Mello, Dieterle and Duckworth 2017; Greene, Caracelli and Graham 1989).

9.4.1 Congruent results

The results of the facial analysis and the self-reports on the effects of multimedia principles on boredom, neutrality, valence, and arousal are largely congruent and consistent with prior expectations, as discussed in Section 9.1. The majority of the applied design principles showed large positive effect on variables indicating emotional states that support learning, as identified in the literature, namely valence, arousal, happy, and all of them were associated with the decrease of the intensity of the state that is considered as

impairing learning, namely, “boredom”, and an expression that might indicate an indifferent state, namely, “neutral”.

9.4.2 Incongruent results

9.4.2.1 Sadness, anger, disgust and self-reported interest

A common feature of the incongruent results was that, for certain multimedia features (such as signalling or multimedia), the automated facial analysis identified negative emotions, which were, in several cases, explicitly contradicted by the interview participants.

The dominance of the expressions “sad” and “angry,” as indicated by FaceReader, may suggest that the viewing experience primarily triggered negative emotions. Retrospective interviews, however, showed that a considerable number of episodes in which automated facial analysis identified an increased level of “sad” were found by students to be particularly engaging and interesting. The following exchange illustrates this clearly:

Interviewer: “The rate of sadness was still high here as well [...] it appeared alongside this graph. Did you not find this part particularly engaging?”
P1: “No, actually, I think I paid really close attention here because it was quite interesting.”
(P1, semi-structured interview)

To further explore the relationship between boredom and sadness, a subsequent analysis was conducted, as described in Section 8.7.3.6, which enabled comparison of the facial analysis values during the periods students identified as interesting and boring. The Wilcoxon signed-rank test indicated no significant difference between sadness levels in conditions retrospectively perceived as boring and interesting ($Z = 0.014$, $p = 0.994$), and the effect size was negligible (rank-biserial correlation $r_{rb} = 0.003$), suggesting no meaningful difference between the two states. Figure 24 shows that changes in sadness (as identified by FaceReader) exhibit high diversity across participants, which may be one reason the emotions reported in the self-report could not be measured using facial analysis tools. Further possible reasons are discussed in Section 10.3.

Table 18: Facial expression identified as sadness, during periods of self-reported interest and boredom

Paired Samples T-Test

Measure 1		Measure 2	W	z	p	Rank-Biserial Correlation	SE Rank-Biserial Correlation	95% CI for Rank-Biserial Correlation	
								Lower	Upper
Sad | SR boredom	-	Sad | SR Interest	391	0.014	0.994	0.003	0.182	-0.343	0.348

Note. Wilcoxon signed-rank test, JASP (Version 0.19.2)

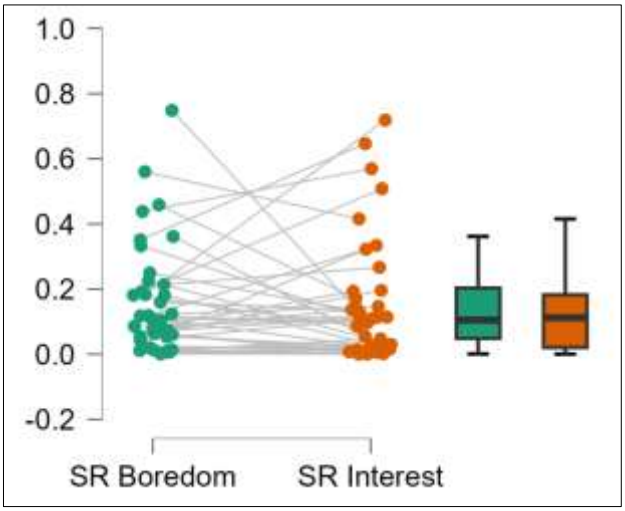


Figure 24: Mean level of sadness, during periods of self-reported interest and boredom

Similarly to sadness, several immediacy-related features showed large effect sizes for lower FaceReader-labelled *anger,* while clarity-related features, especially those reported as engaging, such as "signalling" (r_{rb} = -0.65) and "visual elements" (r_{rb} = -0.468) were associated with higher values of *anger* (see Table 3 of Appendix G). During the retrospective interviews, similarly to segments in which facial expressions indicated high levels of sadness, several segments marked by *elevated anger* by automated facial analysis were described by students as *interesting*.

The facial expression identified as "disgust" was the only negative emotion to show a unique pattern: an increase in its intensity, similar to arousal and valence, was associated with the presence of nearly all multimedia learning features (see Table 19). The underlying reasons for this occurrence are not examined in the present study and warrant further investigation.

Table 19: Effect of certain multimedia principles on "disgusted" as measured by FaceReader, using the Wilcoxon signed rank test (JASP (Version 0.19.2).

Wilcoxon signed rank test (Two-tailed)

Measure 1		Measure 2	W	z	p	Rank	Holm-Bonferroni corrected p	Adjusted significance	Effect Size	SE Effect Size	95% CI for Effect Size	
											Lower	Upper
Disgusted | No visuals	-	Disgusted | Visual element	278.0	-3.470	0.000529	1	0.00582	Significant	-0.564	0.161	-0.743	-0.310
Disgusted | Redundant	-	Disgusted | Not redundant	301.0	-3.248	0.001000	2	0.01000	Significant	-0.528	0.161	-0.719	-0.262
Disgusted | No face	-	Disgusted | Face	330.0	-2.968	0.003	3	0.02700	Significant	-0.482	0.161	-0.688	-0.205
Disgusted | Not Coherent	-	Disgusted | Coherent	382.0	-2.466	0.014	4	0.11200	Non significant	-0.401	0.161	-0.631	-0.106
Disgusted | No Smile	-	Disgusted | Smile	446.0	-1.849	0.065	5	0.45500	Non significant	-0.300	0.161	-0.557	0.009
Disgusted | No personalization	-	Disgusted | Personal	451.0	-1.800	0.073	6	0.43800	Non significant	-0.293	0.161	-0.551	0.017
Disgusted | No signalling	-	Disgusted | Signalling	813.0	1.694	0.091	7	0.45500	Non significant	0.275	0.161	-0.036	0.538
Disgusted | No example	-	Disgusted | Example	494.5	-1.380	0.169	8	0.67600	Non significant	-0.224	0.161	-0.486	0.090
Disgusted | No story	-	Disgusted | Story	515.0	-1.183	0.239	9	0.71700	Non significant	-0.192	0.161	-0.472	0.123
Disgusted | No humour	-	Disgusted | Humour	516.0	-1.173	0.243	10	0.48600	Non significant	-0.191	0.161	-0.471	0.125
Disgusted | Slide rest	-	Disgusted | Slide first minute	540.0	-0.941	0.349	11	0.34900	Non significant	-0.153	0.161	-0.440	0.163

Note. For the Wilcoxon test, the effect size is given by the matched rank biserial correlation.

Note. Shaded cells with grey indicate small (r_{rb} > 0.1), medium (r_{rb} > 0.3) and large (r_{rb} > 0.5) effect sizes, as defined by (Cureton, 1956). The lightest grey indicates non-significant but noteworthy results (Holm-Bonferroni corrected p ≥ 0.05).

9.4.2.2 Signalling and storytelling

Though the majority of the multimedia features showed similar tendencies, two followed different patterns: *Signalling* was the only feature associated with lower valence or arousal values and showed large effect sizes for "happy" (r_{rb} = 0.722), "*angry*" (r_{rb} = -0.65) and "sad" (r_{rb} = -0.645, see Table 2: of Appendix H). *Storytelling* also differed from the general pattern, as no statistically significant differences in positive-emotion indicators were observed between segments with and without storytelling. This deviated from the expected pattern based on both the interview findings and the literature.

9.4.2.3 Comparing FaceReader's boredom data with self-reported boredom and interest

As explained in Section 8.7.3.6, based on the interview data, for each student, two individual 30-second segments of the instructional video stimulus were identified: one segment the student described as the *most interesting* and one segment described as *the most boring*, that is, the segment where maintaining attention was the most challenging. For each participant, the facial analysis data of each participant's most interesting and most boring 30-second segments were compared (see Table 18)[10]. Although these periods differ across students, the t-test analysis allowed comparison of these values within each

[10] Shapiro–Wilk indicated non-normality for all variables except boredom (p = .768); thus, Wilcoxon signed-rank tests were used for non-normal data and Student's *t*-tests for normal data.

participant. (The different data sources were linked only through pseudonymous nicknames, and no personally identifiable information was stored in the database.)

According to the Shapiro-Wilk normality test (see Table 20), these data did not follow a normal distribution, except for the boredom values (p = 0.768); therefore, the Wilcoxon signed-rank test was applied to estimate effect size in these cases, and the Student T-test was used for normally distributed data.

Table 20: Assumption test for normality applied for the calculation of the t-test comparing emotions measured during self-reported boredom and interest.

Test of Normality (Shapiro-Wilk)

			W	p
Boredom | SR Boredom	-	Boredom | SR Interest	0.98	0.768
Valence | SR Boredom	-	Valence | SR Interest	0.92	0.009
Arousal | SR Boredom	-	Arousal | SR Interest	0.92	0.011
Neutral | SR Boredom	-	Neutral | SR Interest	0.93	0.014
Happy | SR Boredom	-	Happy | SR Interest	0.67	< .001
Sad | SR Boredom	-	Sad | SR Interest	0.93	0.013
Angry | SR Boredom	-	Angry | SR Interest	0.45	< .001
Surprised | SR Boredom	-	Surprised | SR Interest	0.7	< .001
Scared | SR Boredom	-	Scared | SR Interest	0.8	< .001
Disgusted | SR Boredom	-	Disgusted | SR Interest	0.47	< .001

Note. Significant results suggest a deviation from normality.

Note. The shaded cell indicates a normal distribution.

Although, as Table 21 shows, none of the measured differences was significant, certain trends can be observed, which are consistent with the self-reported interviews: in the case of the facial expressions of “neutral”, “angry”, “scared” and “boredom” (highlighted in Table 21 in light grey), *moderately strong effect sizes* can be observed between stages identified as boring and interesting across the full duration of these segments. Though these differences are non-significant, as the p-values adjusted using the Holm–Bonferroni correction range from 0.7 to 1, this pattern may suggest that *higher levels of interest co-occur with lower levels of neutral, angry, scared, and bored* states, while higher levels of boredom may be associated with the reverse tendency.

Table 21: Intensity and level of affective states during self-reported boredom and interest periods

Paired Samples T-Test

Measure 1		Measure 2	Test	Statistic	z	df	p	Rank	Holm-Bonferroni corrected p	Adjusted significance	Effect Size	SE Effect Size	95% CI for Effect Size	
													Lower	Upper
Neutral | SR Boredom	-	Neutral | SR Interest	Wilcoxon	523	1.856		0.064	1	0.70	Non-significant	0.341	0.182	-0.005	0.614
Arousal | SR Boredom	-	Arousal | SR Interest	Wilcoxon	269	-1.472		0.143	2	1.00	Non-significant	-0.274	0.184	-0.569	0.083
Angry | SR Boredom	-	Angry | SR Interest	Wilcoxon	491	1.409		0.162	3	1.00	Non-significant	0.259	0.182	-0.095	0.555
Scared | SR Boredom	-	Scared | SR Interest	Wilcoxon	473	1.158		0.253	4	1.00	Non-significant	0.213	0.182	-0.143	0.52
Happy | SR Boredom	-	Happy | SR Interest	Wilcoxon	305	-0.950		0.346	5	1.00	Non-significant	-0.177	0.184	-0.496	0.184
Boredom | SR Boredom	-	Boredom | SR Interest	Student	0.723		38	0.474	6	1.00	Non-significant	0.116	0.112	-0.2	0.43
Valence | SR Boredom	-	Valence | SR Interest	Wilcoxon	361	-0.405		0.694	7	1.00	Non-significant	-0.074	0.182	-0.409	0.278
Disgusted | SR Boredom	-	Disgusted | SR Interest	Wilcoxon	367	-0.321		0.756	8	1.00	Non-significant	-0.059	0.182	-0.396	0.292
Surprised | SR Boredom	-	Surprised | SR Interest	Wilcoxon	383	-0.098		0.929	9	1.00	Non-significant	-0.018	0.182	-0.361	0.329
Sad | SR Boredom	-	Sad | SR Interest	Wilcoxon	391	0.014		0.994	10	1.00	Non-significant	0.003	0.182	-0.343	0.348

Note. For the Student t-test, effect size is given by Cohen's d. For the Wilcoxon test, effect size is given by the matched rank biserial correlation.

9.4.2.4 Correlations between questionnaires and facial analysis

A correlation analysis has been conducted to triangulate the results of automated facial expression analysis and interview outcomes with the results of the questionnaire measuring students' positive and negative affectivity before and following the video watching, as well as their cognitive and emotional interest in the learning material and their perception of the instructor's immediacy, also after watching the video (see Table 22).

- A very strong positive association was found between emotional and cognitive interest, $\rho(49) = .961$, $p < .001$, However, because correlations above .90 may indicate redundancy, and because both values derive from the two subscales of the Student Interest Scale (Mazer 2017), this result likely reflects considerable overlap between the subscales rather than a substantively informative relationship.
- *Positive affectivity and cognitive interest*: Weak but significant positive correlations were found between the increase in positive affectivity and cognitive interest, $\rho(49) = .352$, $p = .013$, suggesting a small but meaningful association.
- *Boredom and positive affectivity*: A significant positive correlation was found between the mean value of individuals' boredom data above their individual 10% peak (for its calculation, see Section 8.7.3.6) and the increase of their positive affectivity (i. e., positive emotional state) after the observation, compared to their level before watching the video. The results, $\rho(49) = .293$, $p = .039$ indicate a weak-to-moderate monotonic relationship. Similar results have been found for the frequency of high-level boredom states, calculated as the number of seconds spent above 90% of the maximal boredom value per participant.

Table 22: Correlation between the facial expression of 'boredom', emotional and cognitive interest, and changes in positive and negative affectivity after video watching.

Spearman's Correlations

			Spearman's rho		p	Effect size (Fisher's z)	SE Effect size
Boredom_meanGreaterThanMax90	-	Increase in negative affectivity	0.155		0.282	0.156	0.146
Boredom_meanGreaterThanMax90	-	Increase in positive affectivity	0.293	*	0.039	0.302	0.148
Boredom_countGreaterThanMax90	-	Increase in positive affectivity	0.297	*	0.037	0.306	0.148
Boredom_countGreaterThanMax90	-	Emotional interest	-0.023		0.875	-0.023	0.146
Increase in negative affectivity	-	Increase in positive affectivity	0.219		0.126	0.223	0.147
Increase in negative affectivity	-	Emotional interest	-0.135		0.355	-0.136	0.147
Increase in negative affectivity	-	cognitive interest	0.004		0.977	0.004	0.146
Increase in negative affectivity	-	Immediacy	-0.019		0.899	-0.019	0.146
Increase in positive affectivity	-	Emotional interest	0.154		0.290	0.155	0.148
Increase in positive affectivity	-	cognitive interest	0.352	*	0.013	0.368	0.150
Increase in positive affectivity	-	Immediacy	0.207		0.154	0.210	0.148
Emotional interest	-	cognitive interest	0.961	***	< .001	1.958	0.166

* p < .05, ** p < .01, *** p < .001

Strong and highly significant negative correlations of neutral with sad ($\rho(50) = -.681$, $p < .001$), and moderate but highly significant and negative correlation with angry ($\rho(50) = -.455$, $p = .001$), and disgusted ($\rho(50) = -.370$, $p = 008$) expressions represent an expected pattern and may be interpreted as an indication of the internal consistency of the facial expression measures rather than as a substantively novel finding.

The findings of the current section are partly in line with Analytic Expectation 3, as both convergence and meaningful discrepancies were observed across methods, as expected; however, the ratio of incongruent results was higher than expected, and two of the design features showed strikingly and markedly different associations with affective patterns than was expected based on the interviews.

10 DISCUSSION

While the previous chapter presented the results in relation to the research questions, the present chapter discusses how these findings relate to the four interconnected research gaps identified in Section 1.2, in line with the theoretical and methodological issues developed in Chapters 3, 4, and 7. Particular attention is given to the discussion of unexpected cross-method discrepancies. Based on comparisons of the different data sets and the literature, Section 10.3 identifies possible explanations for the observed discrepancies.

10.1 Boredom in the context of instructional video viewing

The finding that boredom and interest did not consistently change in opposite directions (see Section 9.2.1) supports the view that boredom cannot be understood simply as the absence of interest. Rather, the two states may coexist or vary in partly independent ways, depending on whether learners can sustain attention and perceive sufficient meaning in the activity. This interpretation is compatible with the MAC model, which conceptualises boredom in terms of unsuccessful attentional engagement and/or insufficient meaning rather than as the direct opposite of interest (Westgate and Wilson 2018). In the present study, this distinction is particularly relevant because some video segments identified as interesting by participants were not necessarily characterised by uniformly low boredom-related facial indicators. Accordingly, interest and boredom may represent partially overlapping rather than mutually exclusive aspects of the learning experience.

The temporal patterns observed around the section most frequently identified as boring (Section 9.2.4.3) add a cautious perspective to the question of whether boredom is necessarily characterised by low arousal. Within the present context, lower arousal co-occurred with the period most frequently identified as boring, but arousal subsequently showed higher values while valence remained below its initial level. This pattern suggests that the FaceReader-derived affective indicators associated with periods of reported boredom may vary over time rather than correspond to a single, stable valence-arousal profile. Earlier distinctions between boredom types characterised by different levels of arousal (Goetz et al. 2014) are compatible with such variability, but the present data do not allow the identification of different boredom types or transitions between them. Rather, the findings suggest that boredom is worth examining as a state that changes over the

course of an activity, rather than only at isolated points in time. How such changes unfold within individual boredom episodes warrants more direct investigation.

Across the analyses, boredom emerged as a changing rather than uniform response during video viewing. Boredom-related indicators varied over time, and the patterns observed after the period of highest boredom did not simply return to those observed beforehand. Boredom also co-occurred with reported effort to sustain attention and did not consistently move in the opposite direction to interest. Together, these findings suggest that boredom during learning cannot be reduced to low stimulation or the absence of interest.

The findings also speak to the question raised in Section 4.3.3 about what makes an activity boring. Boredom co-occurred with participants' reported effort to sustain attention. Thus, within the present context, boredom was not limited to situations of insufficient attentional demand; it could also occur when maintaining attention became difficult. This is consistent with the MAC model's account of attentional mismatch, although the present study cannot establish whether attentional difficulty preceded boredom or followed from it.

10.2 Instructional communication, multimedia features and affective responses

One of the main patterns depicted by the summary heatmap (Figure 23), is that video design *features* promoting *affective engagement* (Bolkan and Griffin 2018; Mazer 2013), which were also identified as immediacy-related strategies (see Section 5.2.1) or relationality-oriented goals (Mottet, Frymier and Beebe 2006), showed stronger positive associations with valence and arousal than clarity-related strategies. Their presence co-occurred with higher intensity of facial expressions coded as "happy" and with lower intensity of most expressions identified as negative or neutral according to the categorisations specified in Ekman's coding system (Ekman and Friesen 1978), including "bored", "sad", "angry", "surprised", and "scared".

Results of the observation and interviews suggest that both immediacy and clarity are important components of teachers' communication. In the examined context, the most effective immediacy behaviour, according to both the facial analysis results and student interviews, was the appearance of teacher's face.

Examination of the co-occurrence of multimedia design features reveals the most consistent pattern: facial expressions identified by automated analysis as neutral and bored, with almost all features associated with a decrease in these states. The observed patterns suggest that segmentation and integrated learning activities may be particularly relevant to sustaining attention in instructional settings.

10.3 Interpreting incongruent findings

The current research used facial expression analysis as the primary strategy to measure affective states and administered self-report questionnaires and semi-structured interviews to support the interpretation of these results and to gain additional insight into the participants' opinions and affective responses to the instructional video. The different data sources provided both congruent and incongruent results. In the latter cases, multimodal data has enabled the establishment of conclusions that helped contextualise and interpret the observed discrepancies or identify directions for further research. It can thus be concluded that the utilisation of multimodal data collection methodologies facilitated the identification of refined, multifaceted findings.

The results for the paired-sample tests for boredom, arousal, valence and "neutral" showed congruent patterns across measurement and analysis methods, as discussed in 0. The positive effect sizes for the increase in "happy" across most video design features were similarly in line with expectations based on the interview results.

Correlation analysis revealed a weak but significant positive correlation between increases in positive affectivity and cognitive interest, $\rho(49) = .352$, $p = .013$, suggesting a small but meaningful association and indicating that experiencing cognitive interest contributes to perceived positive feelings related to the learning experience.

Incongruency across results from different data sources was observed in the increase in "angry", "sad", "disgusted" and "scared" associated with clarity-related video features (see Figure 23).

The incongruencies of the results discussed in Section 9.4.2 may be associated with different possible factors:

1. the weak relationship between mental states and bodily signals;

2. the limited reliability of identifying affective-cognitive changes from facial data;
3. co-occurrence of multiple affective states and overlaps between their expression
4. bias resulting from retrospective evaluation of a situation;
5. data noise related to disturbing factors.

10.3.1 Relationship between mental states and bodily signals

The main reason behind incongruency is, as scholars of the field of affective computing underline, that despite of the sophisticated measurement opportunities, examination of facial expressions and non-verbal behaviour can only be used to *infer* unobservable mental and emotional states, since "there are weak relationships between mental states and bodily signals" (D'Mello, Dieterle and Duckworth 2017, p. 117), thus instead of one-to-one correspondences between emotions and facial expressions or nonverbal behaviours, emotions can follow "prototypical multimodal patterns" (Ekman 1992; Keltner et al. 2019, pp. 139–146).

10.3.2 Reliability of identifying affective-cognitive changes from facial data

The second limitation concerns facial expressions: as indicated by Ekman (1984) and Berka et al. (2007), *emotional changes* are better expressed through facial changes, whereas *cognitive activation* has a clearer and stronger effect on the peripheral nervous system. Immediacy-related behaviours and features were associated with more positive affective indicators, and the corresponding facial-activity patterns were broadly consistent with students' self-reports. Clarity-related strategies may support focus, reflection, attention, cognitive effort and long-term engagement (Bolkan 2017; Bolkan and Griffin 2018; Titsworth et al. 2013); however, they may be associated with negative emotions.

10.3.3 Co-occurrence of multiple affective states and overlaps between their expression

According to the results discussed in Section 9.2.4.1, *boredom* and *interest* showed such wide variability between individuals that, without data normalisation, only a few phases could be identified as triggering boredom or interest, and therefore only a few, very cautious conclusions could be formulated. In most cases, affective states do not occur as distinct experiences but as a mixture of emotions that fluctuate in parallel. As a result, identifying

specific emotions may be misleading, as certain facial actions can be part of combinations that indicate different, and even opposing, emotional states (Ekman 2003).

One incongruent finding concerned signalling, a feature that students evaluated positively in the retrospective interviews but that showed a different pattern in the FaceReader-derived indicators. In particular, signalling was associated with lower valence or arousal values and showed large effect sizes for several FaceReader-labelled expressions, including “happy”, “angry”, and “sad” (see Table 2: of Appendix H). This discrepancy should be interpreted cautiously. As Ekman (Ekman 2003), notes, elements of facial movement may contribute to multiple emotional configurations, and facial expressions do not provide a one-to-one indication of underlying emotional or cognitive states. Accordingly, the FaceReader-labelled “anger” observed during signalling should not necessarily be interpreted as experienced anger; it may reflect facial activity occurring during other affective or cognitive processes.

Relationship between facial expression of *anger* and *focus* or effective learning has been observed in several studies. This may be related to the fact that Action Unit (AU) 4 (brow lowerer) and AU7 (lid tightener) can occur in both anger and focused concentration; however, differences in their intensity and their combinations with other AUs may influence whether the resulting expression is interpreted as anger or as focused concentration (Ahn and Harley 2020; Ekman and Friesen 1978; Grafsgaard 2014).

One of the surprising results of facial analysis was the high intensity of *sadness* in the participants' facial expressions, while according to the interviews, a considerable part of these episodes was assessed as particularly engaging and interesting by students. With regard to *sadness*, several studies have drawn attention to the fact that the changes of facial actions are very similar in the case of sadness and interest (Hetland et al. 2019): “Both interest and sadness are characterised by activation in AU1 (inner brow raise) and AU4 (brow lowered). The only difference is found in mouth gestures, where interest is described by activation of AU24 (lip pressure), and sadness is expressed by activation of AU15 (lip corner puller)” (Hetland et al. 2019, p. 4). Similar effects are discussed by Reschly and Christenson (2022) when they highlight that engagement and disengagement can occur simultaneously, and by further studies concluding that the same emotion (such as smile,

frustration, confusion) can contribute both to engagement and disengagement, depending on length and context (Arguel et al. 2019; D'Mello et al. 2014).

10.3.4 Negative affectivity during an activity is not equal to disengagement

Recall bias in retrospective evaluation. A study comparing extreme-sport athletes' emotions before, during, and after a downhill bike descent found incongruence between their retrospectively reported emotions (happy, interested) and their in-the-moment facial expression analysis, showing low levels of happiness and interest and high levels of anger and sadness (Hetland, 2019). The study found that the incongruency arose from differences between moment-to-moment facial expressions and retrospective self-reports, rather than from overlapping activation of action units.

Situational confounding. Several studies indicate that negative emotions experienced and expressed during learning can be due to disruptive factors, such as noise and distractions. In these cases, "frustration would not be an indication of disengagement but a signal of attempted engagement being thwarted by the environmental mischief of others" (Betts 2022, p. 653). In the current study, participants reported frustration with technical issues, such as the video player stuttering and a screen size that was too small due to software limitations.

Interest does not necessarily imply positive affect. A further explanation appears particularly consistent with the present findings. The apparent incongruence was not simply due to a general mismatch between facial indicators and retrospective reports. During the interviews conducted immediately after viewing, participants were shown specific segments in which FaceReader indicated relatively high levels of boredom or negative emotions, and they were asked about their experiences at those moments. In many cases, participants did not report boredom; instead, they described the same segments as interesting or engaging. This immediate segment-specific questioning reduces, although does not eliminate, the likelihood that the discrepancy can be explained by retrospective recall alone.

More importantly, a similar pattern emerged independently in the *feature-level analyses*. Several instructional features that participants explicitly identified in the interviews as particularly interesting or appealing were also associated with stronger negative facial

indicators. These included signalling and visual representations, as well as segments in which the instructor introduced new information requiring greater cognitive processing. Thus, negative facial affect did not occur randomly in relation to students' reported interest: it appeared particularly during instructional features that students themselves valued and described as interesting.

Westgate and Wilson's (2018) distinction between interest and enjoyment provides a plausible theoretical interpretation of this pattern. Interest is promoted by novelty and complexity and typically requires greater cognitive effort, whereas enjoyment is more strongly associated with familiar, easily processed and immediately rewarding experiences. Importantly, conditions that increase interest can decrease enjoyment. An instructional experience may therefore be interesting and cognitively engaging without being immediately pleasant. Although recall effects, technical disturbances, and measurement limitations may explain individual instances of incongruence, they appear less able to account for this broader, recurring pattern. The convergence of immediate interview accounts and feature-level findings suggests that the interest–enjoyment distinction provides a plausible interpretation of the present pattern, although it cannot be established as its cause.

10.4 RQ3: Alignment across data sources and methods

This section addresses RQ3 and, drawing on the literature and the empirical findings, discusses potential explanations for the incongruencies observed across the data collected and analysed using different methods.

10.5 Implications for practice

10.5.1 Teaching practice

The results of the current measures were in line with the main implications of the Emotional Response Theory on teaching practice, namely that the success of the teacher-student relationship does not depend on "what teacher does": the decisive factor of the process is "how students respond emotionally" (Mottet, Frymier and Beebe 2006, p. 264).

This means that observations and measurements cannot promise uniform recipes for success; rather, they help instructors interpret students' emotional responses to identify key factors of students' trust, sense of control, and the value of their learning.

The findings, in line with the above conclusions, suggest that the *frequent* and *varied* use of teacher immediacy-, clarity-, and multimedia-oriented strategies can effectively prevent and reduce boredom and neutral emotional states while supporting arousal, positive valence, and engagement, which is especially important given that students differ in their cultural backgrounds, personalities, and responsiveness to particular teaching methods.

The results align with existing research that identifies affective responses as central to engagement in learning, including in higher education. They therefore support the significance of immediacy-related communication behaviours. Though teaching through instructional videos reduces the scope of applicable immediacy behaviours, instructor's face, humour, smile and the use of everyday, familiar, first- and second-person talk are key to increasing students' cognitive and affective engagement.

Despite the importance of relationality, *teacher clarity is equally important*; when there is a strong rapport between students and instructors, clarity increases students' emotional interest as well (Mazer, 2012), and people are more interested when the content seems more understandable (Silvia 2006). It is important to be aware that students have different relational and academic needs: depending on their personalities, age, course context or current situation, these needs can vary. For example, younger students might have higher relational needs, while more mature students may value clarity more than immediacy (Mottet el al., 2006).

In the majority of cases, students recalled instances when their faces appeared to reflect negative emotions, as periods when they experienced attention, cognitive effort, or interest. Taking into account the reasons discussed in Section 10.3.4, this may encourage educators to recognise that negative facial expressions – and indeed, the transient appearance of negative emotions – are often associated with attention, interest and cognitive effort. The present study hypothesises that *it is not negative emotions* or facial expressions that merely suggest them, *but rather boredom and neutral facial expressions that pose a threat to student attention*.

10.5.2 Instructional videos

As Chapter 6 highlighted, in line with the multimedia learning review recommendations (Fyfield, Henderson and Phillips 2022), the current thesis sought to identify which multimedia learning principles and instructional strategies support learning. In the context of the current study, certain features showed larger effects, such as

- embodiment,
- humour,
- personalisation,
- signalling and
- multimedia.

Earlier research established that *embodiment* (i.e., showing the instructor's face and gestures) in instructional videos does not clearly improve learning, although it may influence attention and social presence (Fiorella 2021; Fyfield, Henderson and Phillips 2022). The current study provides context-specific empirical evidence that the appearance of the teacher's face in the video was associated with the highest levels of valence and arousal, as well as the lowest values of boredom. However, student interviews suggested that the effectiveness of this feature – and potentially that of other immediacy strategies – depends on students' perceptions of the teacher's likability. Although the present study did not directly measure teacher enthusiasm, the strong boredom-reducing effect of the teacher's face is consistent with previous findings suggesting that visible expressions of teacher enthusiasm may foster enjoyment and thereby reduce boredom (Bieg et al. 2022; Frenzel et al. 2018).

Humour was a variable mentioned by most students as a behaviour fostering a joyful classroom atmosphere and a healthy connection between students and teachers, but research results show that it needs to be appropriate (i.e., content-relevant, non-offensive, non-stereotypical) and adequately employed (Bolkan, Griffin and Goodboy 2018) to effectively contribute to successful learning (Wanzer and Frymier 1999; Mazer, 2022).

Wide range of features. The segment labelled by students as the "most boring" (Segment 6, see Section 8.3) clearly demonstrates an important takeaway: the effectiveness of instructional videos cannot be adequately understood by focusing only on the features

considered most important in isolation; rather, *multiple factors* need to be taken into account together, and a *wide range of multimedia design features should be applied*. Although this segment included signalling, examples, personalisation, and humour, participants nevertheless reported difficulties sustaining attention during the prolonged, unsegmented eight-minute section. Despite the varied methods used in the examined video and the relevance of its topic, its length of 22.4 minutes proved challenging for students to follow. Previous research on instructional videos suggests that shorter videos are often associated with better attention and engagement, although the optimal length depends on content, learner characteristics, instructional purpose, and the organisation of the material. In the present study, several participants perceived the 22-minute video as too long, and higher boredom indicators were observed during prolonged segments.

Segmenting. As shown by Table 10, the instructional video displayed a gradual build-up of attention-catching elements towards the end: features such as stories, jokes, graphs, and figures became increasingly frequent and prominent in the later sections, whereas more traditional elements, such as text-based slides and slides with few visual elements, were more characteristic of the first half of the video. Nevertheless, boredom indicators gradually increased over time, while valence continued to decline and did not return to its initial level even after periods of high boredom (see Section 9.2.4.3). Periods of elevated boredom may occur even when a video contains multiple attention-supporting instructional features. The present findings therefore suggest that instructional design should aim not only to attract attention but also to limit the duration of segments in which boredom is likely to accumulate. Periods of elevated boredom may occur even when a video contains multiple attention-supporting instructional features. The present findings therefore suggest that instructional design should aim not only to attract attention but also to limit the duration of segments in which boredom is likely to accumulate (Fyfield, Henderson and Phillips 2022; Woolfitt 2015).

Multimedia learning features associated with boredom

RQ1 addressed associations between design features and boredom. The findings indicate that nearly all of the identified design strategies were associated with lower values of FaceReader-labelled “boredom” and “neutral”, while most were also associated with higher arousal and valence values. In addition to this short answer, the current section

summarises the research findings, exploring and confirming the characteristics of boredom and factors that can contribute to its prevention and reduction.

Characteristics of boredom in the context of the current research

Boredom negatively affects students' learning engagement and achievement. Interview findings indicated that periods experienced as boring were associated with reduced attention and greater difficulty recalling the content or details of those episodes. The present design does not establish the direction of this relationship, but the pattern is consistent with previous research linking boredom with reduced attention and learning performance (Daschmann, Goetz and Stupnisky 2014; Pekrun et al. 2014; Sharp et al. 2018).

In the present context, the observed temporal patterns suggest that re-engagement after periods of high boredom may be more difficult than maintaining engagement before such periods arise. The repeated-measures analyses (Section 9.2.4.3) showed significant temporal variation in arousal across the periods surrounding the high-boredom phase, whereas the corresponding variation in valence did not reach statistical significance. Descriptively, arousal decreased during the high-boredom period and subsequently returned towards its earlier level, while valence remained below its initial level. This pattern should therefore be interpreted with caution, but it may suggest that different dimensions of affect do not necessarily recover at the same rate following periods of high boredom. (Eastwood et al. 2012; Pekrun et al. 2014; Westgate and Wilson 2018). The analysis of FaceReader-derived boredom showed considerable individual variation (Section 9.2.4.4). Although boredom differed significantly across the three periods overall, the clearest change was the decrease following the boring episode, while individual trajectories varied substantially. The cluster analysis similarly distinguished participants mainly by their overall level of FaceReader-derived boredom rather than by distinct temporal patterns. Together, these findings support a person–situation interpretation of boredom, suggesting that the same situational context may be associated with different boredom-related responses across learners.

Boredom may be triggered or strengthened by several interconnected factors, such as cognitive overload, underchallenge, lack of meaningfulness or isolation. As discussed in Section 4.3 and summarised in Figure 7, boredom shows individual- and context-specific

variation in its triggers. The findings of the present study similarly suggest substantial individual variation in boredom-related responses, potentially depending on how meaningful or interesting participants perceived the video-viewing task, whether they experienced it as too easy or too difficult, their relationship with the course and the instructor, individual differences in boredom proneness, and contextual factors such as current fatigue.

The findings also speak to the question raised in Section 4.3.3 concerning the conditions under which boredom may emerge. Within the present context, boredom co-occurred with reported effort to sustain attention, suggesting that boredom may arise when maintaining attentional engagement becomes excessively demanding. This pattern is consistent with the MAC model's account of attentional mismatch and adds context-specific evidence to the broader theoretical debate about the role of cognitive demand in the emergence of boredom.

In the instructional video examined in the present study, *boredom indicators tended to increase as viewing progressed.* At the same time, the duration of low-boredom phases decreased, despite the more frequent use of attention-supporting features in later sections of the video.

Instructional features associated with boredom levels

Based on the characteristics presented and the results discussed in the present thesis, the following findings can be identified:

Different instructional features showed different patterns of association with boredom-related indicators. Potential antecedents of boredom may vary across and within participants; moreover, they can co-occur, depending on the specific situation. Several features of teacher communication, such as "avoiding redundancy," "coherence," or "segmenting," were associated with lower boredom indicators and may be particularly relevant in contexts involving underchallenge. Because boredom may arise from different and potentially co-occurring antecedents, no single instructional strategy can be expected to address all instances of boredom. In the present study, a broad range of communication and multimedia design strategies co-occurred with lower boredom indicators across different segments of the video. E.g., participant interviews indicated that students

perceived signalling and visual elements as helpful in maintaining or redirecting their attention when the content was experienced as either challenging or overly obvious. This suggests that, in instructional videos, perceived cognitive demands may be associated not only with overchallenge but also with the monotony of video viewing, a learning format lacking interaction and active learning.

The findings suggest that combining varied immediacy- and clarity-related strategies may be more promising than relying on a single instructional feature, particularly because students differ in their attentional, relational, and academic needs. The lower "boredom" and "neutral" values observed during segments containing a broad range of design features were accompanied by higher arousal and valence indicators. These converging patterns suggest that varied instructional strategies may be associated with sustained affective activation and lower boredom during video viewing.

This interpretation is consistent with the higher arousal and valence values observed during video segments containing these design features, since boredom is usually characterised by the simultaneously low levels of arousal and valence (Goetz, Hall and Krannich 2019b). In line with Mazer's (2022) argument, given the diversity of students' cultural backgrounds and personalities, different methods can be effective in engaging the largest possible proportion of students.

The teacher's face (embodiment) in instructional videos can function as a socially meaningful visual cue for attracting students' attention. The feature that showed the strongest association with lower boredom indicators in the analyses presented in this thesis, and was also strongly supported by the interview findings, was the appearance of the instructor's face (the multimedia principle of *embodiment*). This aligns with earlier research on the distinctive role of the human face in attracting attention from a very early age (Goren, Sarty and Wu 1975; Haxby, Hoffman and Gobbini 2000; Langton, Watt and Bruce 2000). This association may be particularly relevant in online learning contexts, where the need for connection may be heightened, especially when embodiment is combined with personal and second-person-phrased messages, which create a sense of connection that energises the learning process (Matthews, Lodge and Johnstone 2024). However, an important limitation should be acknowledged: although 63% of participants explicitly identified this feature as supporting their attention, the analysed video used it for

only brief segments lasting 10–90 seconds. Research on instructional video suggests that prolonged use of this tool may not necessarily lead to greater attention (Fyfield, Henderson and Phillips 2022).

The observed patterns suggest that *segmenting and integrated learning activities may be particularly relevant to sustaining attention in instructional settings*. According to the relative frequency analysis, the longest boredom-free segment corresponded to the first two minutes of the same table, and its end was identified as the most boring segment. Thus, the subsequently high boredom indicators co-occurred with a prolonged section lacking segmentation. The implementation of a pedagogical strategy that divides a session into smaller segments, preferably interspersed with learning activities, may therefore be relevant to addressing this issue, as highlighted by the principle of integrated learning activities (Fyfield, Henderson and Phillips 2022; Mayer and Fiorella 2021).

RQ2: Co-occurrence between design features and affective response patterns

Findings on affective patterns derived from facial expression data and their co-occurrence with multimedia design features and teacher behaviour are formulated based on the results of the paired-samples tests and summarised in the heatmap in Section 9.3.2. These results suggest that *immediacy-related features co-occurred with indicators associated with more positive and activating affective states.* The heatmap shows clearly distinguishable patterns for immediacy- and clarity-related tools. Features intended to support teacher immediacy, especially face, humour, personalisation, and smile, showed significant associations with higher valence, arousal, and intensity of facial expressions labelled “happy”. Several of these associations were characterised by large effect sizes. These indicators represent positive, and activating affective states associated with engagement (Pekrun et al. 2023). *Immediacy-related features were also significantly associated with lower values of several negative and deactivating affective indicators*, particularly boredom, sadness, and anger.

Several clarity-related features, including visual elements, absence of redundancy, coherence, and examples, *were associated with higher values of arousal, valence*, and “happy”, although the strength and statistical significance of these associations varied across features. Segmenting and signalling did not follow this general pattern. These indicators have been associated with learning-supportive affective states in previous research (Pekrun, Marsh, Suessenbach, et al. 2023). Clarity-related features were also

associated with higher values of several expressions labelled by FaceReader as negative, particularly sadness, anger, disgust, and fear. This result contradicted prior expectations and self-report indicators. Section 1.3 lists possible reasons for this discrepancy in relation to RQ3.

10.5.3 Implications for other fields

Instructional videos are increasingly applied in *organisational knowledge sharing* (Huggett 2021) as they effectively support onboarding, increase the flexibility of the pieces of training, allow participants to progress at their own pace, and facilitate learning for diverse groups (Nederveld & Berge, 2015). However, learning through instructional videos poses similar challenges for recipients in the workplace as it does in the education system: learners experience a decline in engagement with digital learning compared to before, and 53% of learning and development professionals consider their organisation's virtual training experiences ineffective. The effectiveness of instructional videos thus largely depends on their ability to maintain viewers' attention, and this study's findings support in-company knowledge exchange in improving instructional videos by exploring the factors affecting viewers' (dis)engagement.

The present findings are also relevant to other forms of explanatory and instructional audiovisual communication:

- *Public-facing videos* often address heterogeneous audiences and therefore face similar challenges of maintaining attention, structuring information clearly and supporting engagement (Moussiades et al., 2019).
- The findings may also inform software demonstrations, digital tutorials and *user-support materials*, where video design needs to combine clear explanation with efficient pacing and revision based on user response (Moussiades et al., 2019).
- Beyond knowledge transmission, audiovisual communication can also support motivation, interest, attitude formation and a sense of connection with the communicator (Koumi, 2006). These functions are particularly relevant in mediated expert–user communication, such as *professional training, health education* and other explanatory contexts where immediate interpersonal interaction is limited or absent. The present findings on boredom, attention and

relational communication suggest that human presence and relational cues may remain important even under these conditions.

Designed relationality may also be relevant *in mediated expert–user contexts* where immediate reciprocity is limited or absent. Human presence, tone, direct address, facial expression and visual embodiment may help sustain engagement even when the recipient cannot respond directly to the communicator.

10.6 Limitations and directions for future research

The relatively small sample size (N = 55) should be taken into account when interpreting the findings, though this limitation is partly offset by the substantial amount of data analysed from multiple sources, including 73,920 video frames, 2,450 questionnaire responses, and 275 minutes of interview material. The gender distribution of the current sample does not allow for drawing conclusions from differences between male and female results due to the low rate of male participants. The topic of the video presentation in the case study was a particularly popular theme among the participants (brand destruction). As the field of study may impact students' affective and cognitive interests (Brom et al., 2017), to enhance the generalisability of the findings, further steps and directions include replicating the observation, with a larger sample and including instructors teaching a wider range of subjects across different study areas.

The study was conducted under laboratory-like conditions that differed from everyday learning with instructional videos. Students mentioned feeling this difference when they could not take notes or had to avoid typical behaviours, such as checking their phones. Conformity bias can be especially strong when observing student emotions, given the more vulnerable situation and environment (Lewinski, Fransen and Tan 2014). To mitigate it, several studies recommend conducting similar measurements in less controlled environments (i.e., classrooms or students' homes) as well (D'Mello, Dieterle and Duckworth 2017; Zhong *et al.* 2017; Landmann 2023). Although participants were left alone during video viewing to minimise potential observer effects, the possibility of observation-related bias cannot be entirely excluded. Awareness of being recorded and participating in a research setting may still have influenced participants' behaviour or facial expressions. At the same time, this limitation is difficult to eliminate in studies that

require video-based facial analysis, and the adopted procedure was an attempt to reduce researcher presence while maintaining the conditions necessary for data collection.

The application of automated facial recognition systems in educational environments is still in its early stages: algorithms should be trained to recognise learning-specific emotions (Hirt et al., 2018). Additionally, several scholars raised concerns related to the identification of discrete emotions of studies trying to match emotion “labels” with certain features of a stimulus, warning that emotions most often appear to be a mixture of feelings that exist on a broad spectrum beyond the basic emotions, often occurring simultaneously, and they may signal not only the effects triggered by the stimulus but intentions, reflections on the situation or a person involved in it, or personality-related individual feelings (Keltner 2016). Individual differences in emotion regulation can distort emotion expression, intensifying it on some people's faces while suppressing it in others.

Incongruent results between facial expressions indicating negative emotional states and self-reports may align with D’Mello et al.'s (2017) argument: since engagement is a multicomponential construct, different modalities may be better suited to capturing distinct components, with eye gaze and central physiological measures being particularly informative for cognitive engagement. Therefore, clearer insights could be gained by complementing facial analysis with data on gaze direction, head pose, eye closure rate, posture, and hand resting on the head; these indicators can be captured through semi-automated coding of participants’ video recordings. A third approach would be to combine facial analysis with physiological measures and self-report data, which may improve the adequacy of observations aimed at examining academic emotions.

A methodological consideration regarding questionnaires is that self-reports are only very limitedly suitable for measuring emotions, especially for detecting boredom (Kim et al., 2018; Merrifield & Danckert, 2014; Raffaelli et al., 2018).

A further limitation concerns the level of inference that the research design supports. The empirical study examined one instructional video presented by one instructor within one course and institutional context. Although within-participant comparisons provide evidence of systematic associations between specific video features and measured response patterns, these contrasts do not experimentally isolate the features from other stimulus characteristics and therefore do not establish independent causal effects. The

findings should consequently be interpreted as context-bound empirical patterns that can inform theoretical propositions and subsequent experimental or replication studies rather than as estimates of universally applicable instructional effects.

A related limitation concerns the ontological status of automated emotion categories. FaceReader classifications are model-generated indicators based on facial configurations and should not be treated as direct observations of subjective emotional states. The relationship between facial movements and experienced emotion varies across individuals and contexts, and similar facial configurations may occur during different affective or cognitive processes (Barrett et al. 2019). This limitation is particularly relevant to the present study because several expressions labelled as negative emotions co-occurred with periods that participants retrospectively described as interesting or attention-demanding.

11 CONCLUSION

This concluding chapter synthesises the main findings of the thesis and considers their significance in relation to the broader theoretical and methodological problems that gave rise to the study. Particular attention is given to three contributions broadly aligned with the research questions: the interpretation of boredom in learning, the conceptual connection between multimedia learning and instructional communication, and the methodological implications of combining automated facial analysis with self-report data.

11.1 Summary of aims and key findings

In line with the methodological recommendations of academic emotions research, the present study used a multimodal approach, based on results from three measurement methods: self-report questionnaire, automated facial analysis, and semi-structured interviews, to assess students' emotional reactions to the various features of an instructional video. The results supported the assumption that only certain methods are suitable for revealing specific information. Though several emotions identified by facial expression analysis did not converge with self-reports, their changes over time can indicate learning-relevant cognitive processes. Thus, all three measurement methods revealed associations that would not otherwise have become apparent.

A comparison of automated and self-reported measures suggested that, in certain cases, these emotions indicate affective states that hinder learning, but several expressions labelled as negative, such as disgust, contempt, and sadness, may also indicate cognitive processes, such as attention, interest, focus, reflection or mental effort, that are associated with effective learning.

Results of the observation and interviews suggest that both immediacy and clarity are important components of teachers' communication. In the examined context, the immediacy-related feature showing the strongest and most consistent association in this study, according to both the facial analysis results and student interviews, was the appearance of the *teacher's face*. The use of everyday language, personalised messages, and humorous phrases was also associated with more positive and activating affective indicators and was positively evaluated in the interviews. Student interviews also

identified the clarity, relevance, and structure of the information as features associated with their reported interest in the content.

Across the analyses, the observed patterns suggest that, within the present context, re-engagement following periods of high boredom may be more difficult than maintaining engagement before such periods emerge. In the present context, segmented sections were associated with lower boredom indicators. This involves not only dividing instructional videos into shorter segments but also ensuring that slides change every 1–2 minutes, including, e.g., text animation, and, most importantly, that integrated learning activities between shorter segments provide opportunities for active learning.

The application of measurement opportunities that enable automated detection of physiological or affective processes is particularly important in studies of boredom, given the limited and biased perception of boredom during learning. Examination of the co-occurrence of multimedia design features reveals the most consistent pattern: FaceReader-labelled “neutral” and “boredom” indicators, which generally showed lower values during segments containing the examined instructional features.

Student engagement can be examined across different dimensions and at different levels of the educational system. The present study focused primarily on affective engagement at the micro level, addressing only a limited part of the broader set of factors that shape learning. Nevertheless, this level is of particular relevance because communication principles formulated at the instructional interface may remain applicable across differing institutional, regulatory, and economic contexts, provided that instructors are sufficiently responsive to students’ academic and relational needs.

From an instructional communication perspective, the findings do not support a uniform set of prescriptions, but they highlight the importance of considering teachers’ rhetorical and relational goals alongside students’ academic and relational needs. Within the present context, clarity- and immediacy-related features were associated with patterns relevant to attention, interest and affective experience, suggesting that rhetorical and relational communication should be treated as complementary rather than competing orientations. Communication that supports comprehensibility and perceived control may help sustain cognitive involvement, while immediate communication that strengthens instructor presence may support approach-oriented responses towards the content and the learning

situation. These processes may, in turn, provide resources for the effort and self-regulation required when engagement becomes difficult to maintain.

Boredom is especially relevant in this respect because it need not be understood simply as a state to be eliminated. As an aversive signal of unsuccessful engagement, it may indicate to both learner and instructor that attention, meaning, task demands, or the current way of engaging with the activity requires adjustment. The communicative challenge is therefore not only to minimise boredom, but also to create conditions in which learners can reorient attention, recognise value, and re-engage with the learning process. In this sense, clarity can support understanding and a sense of control, while relational communication may contribute to trust, psychological closeness and the willingness to remain in contact with a demanding learning situation rather than withdraw from it.

11.2 Main contributions and novel insights of the thesis

11.2.1 Theoretical contribution to understanding boredom

The findings add to the understanding of boredom in three related ways: they show its variation over time, connect it with difficulty in sustaining attention, and question a simple opposition between boredom and interest.

First, boredom did not appear as a uniform state across the video. Boredom-related indicators changed over time, low-boredom phases became shorter as viewing progressed, and the pattern following the phase of highest boredom differed from that observed beforehand. These findings do not establish a general trajectory of boredom, but they support examining boredom as a process that unfolds over time.

Second, boredom co-occurred with reported effort to sustain attention. This matters because boredom is often associated with too little stimulation, whereas the present findings show that it may also occur when sustaining attention becomes difficult. This pattern fits the MAC model's account of attentional mismatch: boredom can arise when attentional demands are poorly matched to the resources available for engaging with them. The present design cannot determine whether attentional difficulty led to boredom or resulted from it, but the finding broadens the conditions under which boredom needs to be considered.

Third, the findings question the idea that interest is simply the opposite of boredom. Boredom and interest did not consistently move in opposite directions, and moments described as interesting were not necessarily accompanied by uniformly positive facial affect. This distinction becomes clearer when interest and enjoyment are separated. As Westgate and Wilson (2018) argue, the two represent different routes out of boredom: interest concerns engagement with the activity, whereas enjoyment concerns its affective quality. The present findings are consistent with this distinction and suggest that becoming interested and feeling better should not be treated as the same response to boredom.

At a broader level, the literature review and the present findings also point to the need to keep boredom conceptually distinct from engagement. Engagement is a much broader, multidimensional construct, encompassing behavioural, cognitive, emotional, and other components, whereas boredom is a more specific affective-cognitive state. For this reason, the present study did not treat engagement as an outcome that could be directly inferred from boredom or from individual instructional features. Instead, it focused on more proximal indicators, such as attention, interest, affective responses, and learning-related experiences. This distinction also means that boredom should not be treated as the opposite of engagement or as synonymous with disengagement, but rather as one possible process through which engagement may become more difficult to sustain.

This finding also returns to the broader question raised in Section 1.2 about the role of unpleasant affective states in learning. Within the present context, reported interest and cognitive involvement were not always accompanied by more positive FaceReader-labelled expressions. This is consistent with research showing that temporary confusion, frustration, and other unpleasant states can occur during engaged learning, particularly when they do not persist over time (Baker et al. 2010; D'Mello et al. 2014; Arguel et al. 2019). The findings therefore suggest that engaged learning cannot always be characterised by uniformly positive affective responses, and that the role of unpleasant states may depend on their duration and the learning context.

11.2.2 Conceptual contribution: connecting multimedia learning and instructional communication

The distinction discussed in Section 11.2.1 is especially relevant to the present findings. Several instructional features that students explicitly valued and described as interesting

showed significant associations and large effect sizes for higher values of FaceReader-labelled negative expressions. This may indicate that cognitively involving or interesting learning experiences are not necessarily accompanied by more positive facial affect. Within the present context, bringing Westgate and Wilson's distinction between interest and enjoyment into relation with these findings suggests that instructional communication may support engagement through different routes, some more closely related to attention, meaning, and interest, and others potentially more closely related to positive affect or enjoyment. This interpretation does not imply a direct correspondence between the interest–enjoyment distinction and the rhetorical–relational framework, but it offers a novel way of considering why instructional features that appear valuable or engaging may nevertheless coincide with FaceReader-labelled negative expressions.

A further contribution of the thesis is the development of a two-level conceptual framework that directly addresses the conceptual and operational gaps identified in Section 1.2. Multimedia learning offers relatively specific and observable design categories, but its established principles give comparatively limited attention to relational functions of teacher communication. Instructional communication, in contrast, provides a rich vocabulary for rhetorical and relational processes, including clarity and immediacy, but does not offer a single integrated feature-level system that can be readily used to code instructional materials.

The framework developed in this thesis connects these two strengths. At the broader level, the four-dimensional model presented in Figure 8 organises concepts from instructional communication, educational psychology and multimedia learning according to their predominant rhetorical and relational orientations. At the feature level, the conceptual mapping in Table 7 links selected multimedia-learning principles and coded instructional-video features with rhetorical and relational communication functions, using clarity and immediacy as complementary rather than mutually exclusive interpretive lenses. In this way, the framework provides a way to analyse observable instructional-video features while also retaining their broader communicative functions, thereby addressing the gap identified in Section 1.2 between feature-level coding and the conceptual richness of instructional communication.

11.2.3 Methodological contribution

One of the main contributions of the current study may be to illustrate that multimodal measurement can provide access to different layers of students' learning experiences, with facial analysis offering a fine-grained insight into moment-to-moment affective change. Although the analytical and interpretative methodology for data derived from automated facial analysis as a measurement tool is still evolving, the experiences and insights gained support the claim that facial analysis makes a valuable contribution to emotion measurement by enabling fine-grained examination of dynamic emotional changes in learning contexts.

The methodological contribution is also related to the scale and temporal granularity of the facial-expression dataset, which required the development of the staged analytical workflow described in Section 8.7.3, extending beyond software-generated summaries and overall mean values.

A more specific methodological contribution of the thesis lies in the development and iterative refinement of a *segment-based analytical procedure* for handling the large, temporally fine-grained datasets generated by automated facial analysis. The development of this analytical workflow was informed by Landmann's (2023) methodological review of Noldus FaceReader applications and by a separate targeted review conducted for this thesis, which examined how previous studies processed and analysed FaceReader data across different research contexts. Rather than relying solely on FaceReader's graphical outputs or overall mean values, facial-expression data were *temporally aligned with the second-by-second coding* of the instructional video. This made it possible to compare, within the same participant, affective indicators recorded during segments in which a particular instructional or communication feature was present with those recorded during segments in which that feature was absent. Depending on distributional assumptions, these contrasts were analysed using paired-samples Student's *t*-tests or Wilcoxon signed-rank tests, while the risk of false-positive findings arising from multiple comparisons was controlled using Holm–Bonferroni adjustment.

The procedure, therefore, combined fine-grained temporal alignment, within-participant comparison, and corrected inferential testing, thereby transforming highly individual and continuously changing facial-expression data into statistically comparable segment-level

measures. Beyond the present study, this analytical workflow may offer an adaptable approach for future research that uses FaceReader or comparable automated facial analysis tools to examine how short-term affective responses co-occur with precisely timed features of instructional or other dynamic stimuli. Importantly, this procedure was developed iteratively through several analytical stages, with each step informed by methodological insights and limitations identified in prior analyses. Section 8.7.3 documents how this methodology was developed through successive analytical stages. In this process, several analytical approaches identified in the methodological literature and in the targeted review of previous FaceReader studies were applied to the present dataset, enabling an examination of their usefulness and limitations before the final analytical procedure was progressively refined.

The pattern of associations between teaching strategies, behaviours, and learning-related affective indicators did not entirely align with the initial analytical expectations, as the most influential empirical and theoretical studies of academic emotions consistently highlight the effectiveness of positive emotional states (Mazer 2013b; Mottet, Frymier and Beebe 2006; Pekrun, Marsh, Elliot, et al. 2023; Waldbuesser and Titsworth 2022). However, given the characteristics of automated facial analysis, the findings suggest that higher-intensity facial expressions labelled as negative emotions may, in some cases, coincide with learning-related processes such as sustained attention or mental effort, rather than reflecting affective states alone.

The appearance of negative emotions during learning is partly because temporary, frustrating negative feelings, such as confusion, mild anxiety, or fluctuations between positive and negative emotions, are an inherent component of engaged learning, provided they do not persist over time. As previous research has shown (Arguel et al. 2019; Baker et al. 2010; D'Mello et al. 2014), these emotional states do not impair learning as much as boredom and indifference do.

Bodily signs of boredom can be identified by the Facial Action Coding System only to a limited extent (Keltner 2016). Results of the current study suggest that facial actions may be less universal in relation to cognitive processes than to emotional changes; therefore, FaceReader's facial analysis function may be more suitable for detecting emotional changes than cognitive ones. Although these emotional estimates from FaceReader were

partly contradicted by self-reports, fluctuations and peak phases of anger, sadness, and contempt over time can indicate learning-relevant cognitive processes. One possible way to obtain more accurate information about fluctuations in cognitive states is to supplement facial analysis software with learning-specific custom expressions (i.e., customizable combinations of facial movements).

11.3 Concluding remarks

The present study examined how selected instructional video features were associated with affective-cognitive responses during video-based learning, addressing several research gaps and areas of inquiry identified in studies of multimedia learning (Fyfield, Henderson and Phillips 2022; Waldbuesser and Titsworth 2022) and instructional communication.

The findings suggest that immediacy- and clarity-related features were associated with distinct affective-cognitive responses, expressed as explicit verbal reflections and observed as facial expressions. The strongest positive associations were observed for several immediacy-related design cues, particularly the appearance of the teacher's face on the screen. While some relationships converged across methods, others suggested that facial actions labelled as negative emotions may reflect cognitive effort and interest rather than negative attitudes.

One of the clearest and most promising findings of this thesis, supported consistently across multiple data sources, formats, and their interrelations, was that across the examined instructional features, the overall pattern was predominantly associated with lower boredom and neutral indicators, although the strength and statistical robustness of these associations varied substantially across features. This pattern is particularly relevant because boredom represents an important obstacle to learning engagement and is far more detrimental than a mere lack of interest or attention: as a negative and deactivating state, it can generate a vicious cycle of low attention, poor recall, low motivation, and a low level of perceived value and control in the learning process. Given that the emergence of boredom can vary considerably across individuals, the observed patterns suggest potential value in combining varied multimedia design principles, although their relationship with boredom requires further investigation using designs that permit causal inference.

The significance of emotions in cognitive interest and engagement has already been supported by several studies but recent research clearly draws attention to the fact that emotional engagement and relationality play a more central role in student engagement than previously believed (Houser and Hosek 2022; Kahu and Nelson 2018; Matthews, Lodge and Johnstone 2024; Smith and Tinto 2022), and it also indicates the changing needs of the student population over time: They have an increasing expectation that the learning process should be infused and driven by emotional interest (Matthews, Lodge and Johnstone 2024). Nevertheless, these findings do not imply that clarity-oriented instructional interventions are of secondary importance. The effects of strategies that support cognitive and affective engagement cannot be separated: as prior research concludes, both immediacy- and clarity-related instructional interventions enhance students' affective and cognitive engagement.

Both the conceptual mapping's conclusions and the empirical study's results highlighted that, although the need to feel connected may vary considerably among learners, it can become particularly strong in the impersonal, solitary environment of instructional videos.

Although asynchronous instructional videos do not allow genuine transactional interaction within the video itself, this does not necessarily preclude relational communication. The conceptual perspective developed in this thesis suggests that instructors can intentionally incorporate cues of immediacy, personal presence and responsiveness into a structurally one-way medium, creating what may be understood as *designed relationality*. Such relationality cannot substitute for reciprocal teacher–student interaction, and its relevance is likely to vary across learners and contexts. Its value should therefore not be reduced to a fixed set of techniques that are assumed to work equally well for everyone. Rather, its potential may lie in an instructor's willingness to be recognisably present through immediacy, rapport-building, appropriate self-exposure and other need-supportive forms of communication that reflect their own personality, enthusiasm and way of relating to students.

The present findings cannot establish which forms of relational communication will be most meaningful for different learners, but they support the view that asynchronous instructional videos need not be relationally empty. At the same time, this represents only one dimension of student engagement: both instructors and learners require institutional

and broader structural conditions that support the *attentional* and *meaningful* engagement required for learning.

12 REFERENCES

Ahn, B.T. and Harley, J.M. (2020): 'Facial expressions when learning with a Queer History App: Application of the Control Value Theory of Achievement Emotions', *BRITISH JOURNAL OF EDUCATIONAL TECHNOLOGY*, 51(5), pp. 1563–1576. Available at: https://doi.org/10.1111/bjet.12989.

ALLEA - All European Academies (2023): *The European Code of Conduct for Research Integrity*. DE: ALLEA - All European Academies. Available at: https://doi.org/10.26356/ECoC (Accessed: 14 December 2025).

Altuwairqi, K. et al. (2021): 'A new emotion–based affective model to detect student's engagement', *Journal of King Saud University - Computer and Information Sciences*, 33(1), pp. 99–109. Available at: https://doi.org/10.1016/j.jksuci.2018.12.008.

Amini, A. and Weisi, H. (2023): 'The intersectionality of teacher immediacy, emotioncy and willingness to communicate in higher education context', *Journal of Applied Research in Higher Education*, 15(1), pp. 34–50. Available at: https://doi.org/10.1108/JARHE-05-2021-0175.

Andersen, J.F. (1978): *The Relationship Between Teacher Immediacy And Teaching Effectiveness.* West Virginia University. Available at: https://researchrepository.wvu.edu/etd/8405 (Accessed: 13 February 2025).

Arguel, A. et al. (2019): 'Seeking optimal confusion: a review on epistemic emotion management in interactive digital learning environments', *Interactive Learning Environments*, 27(2), pp. 200–210. Available at: https://doi.org/10.1080/10494820.2018.1457544.

Artino Jr., A.R. and Stephens, J.M. (2009): 'Beyond grades in online learning: Adaptive profiles of academic self-regulation among Naval Academy graduates.', *Journal of Advanced Academics*, 20(4), pp. 568–601. Available at: https://doi.org/10.1177/1932202X0902000402.

Asmar, C., Page, S. and Radloff, A. (2015): 'Exploring anomalies in Indigenous student engagement: findings from a national Australian survey of undergraduates', *Higher Education Research & Development*, 34(1), pp. 15–29. Available at: https://doi.org/10.1080/07294360.2014.934334.

Atkinson, J.W. (1957): 'Motivational determinants of risk-taking behavior.', *Psychology Review* [Preprint]. Available at: https://www.semanticscholar.org/paper/Motivational-determinants-of-risk-taking-behavior.-Atkinson/9f9c2c167d230571e3f3f29ed355367f6d588c9e (Accessed: 19 April 2026).

Baker, R.S.J.D. et al. (2010): 'Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive–affective states during interactions with three different computer-based learning environments.', *International Journal of Human-Computer Studies*, 68(4), pp. 223–241. Available at: https://doi.org/10.1016/j.ijhcs.2009.12.003.

Bandura, A. (1977): 'Self-efficacy: Toward a unifying theory of behavioral change.', *Psychological Review*, 84(2), pp. 191–215. Available at: https://doi.org/10.1037/0033-295X.84.2.191.

Barrett, L.F. et al. (2019): 'Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements', *Psychological Science in the Public Interest*, 20(1), pp. 1–68. Available at: https://doi.org/10.1177/1529100619832930.

Bench, S. and Lench, H. (2013): 'On the Function of Boredom', *Behavioral Sciences*, 3(3), pp. 459–472. Available at: https://doi.org/10.3390/bs3030459.

Ben-Eliyahu, A. et al. (2018): 'Investigating the multidimensionality of engagement: Affective, behavioral, and cognitive engagement across science activities and contexts', *Contemporary*

Educational Psychology, 53, pp. 87–105. Available at: https://doi.org/10.1016/j.cedpsych.2018.01.002.

Berlyne, D.E. (1960): *Conflict, arousal, and curiosity.* New York: McGraw-Hill Book Company. Available at: https://doi.org/10.1037/11164-000.

Betts, J. (2022): 'Measuring Student Engagement: New Approaches and Issues', in A.L. Reschly and S.L. Christenson (eds.) *Handbook of Research on Student Engagement*. Cham: Springer International Publishing, pp. 647–658. Available at: https://doi.org/10.1007/978-3-031-07853-8_32.

Bieg, S. et al. (2022): 'Teachers' enthusiasm and humor and its' lagged relationships with students' enjoyment and boredom - A latent trait-state-approach', *Learning and Instruction*, 81, p. 101579. Available at: https://doi.org/10.1016/j.learninstruc.2021.101579.

Bieleke, M., Wolff, W. and Martarelli, C.S. (2024a): 'Boredom is not boring', *The Routledge International Handbook of Boredom*. Routledge.

Bieleke, M., Wolff, W. and Martarelli, C.S. (eds.) (2024b): *The Routledge international handbook of boredom*. Abingdon, Oxon: Routledge.

Biggs, J.B. and Tang, C.S. (2011): *Teaching for quality learning at university: what the student does*. 4th ed. Maidenhead: McGraw-Hill/Society for Research into Higher Education/Open University Press.

Blake, S., Capper, G. and Jackson, A. (2022): *Building Belonging in Higher Education Recommendations for developing an integrated institutional approach*. Wonkhe. Available at: https://wonkhe.com/wp-content/wonkhe-uploads/2022/10/Building-Belonging-October-2022.pdf.

Bloom, B.S. (ed.) (1956): *Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive Domain*. New York.

Boehme, K.L. et al. (2025): 'Is Boredom the Opposite of Interest? A Longitudinal Reciprocal Effect Study', *Educational Psychology Review*, 37(1), p. 15. Available at: https://doi.org/10.1007/s10648-025-09991-5.

Bolkan, S. (2017): 'Development and validation of the clarity indicators scale', *Communication Education*, 66(1), pp. 19–36. Available at: https://doi.org/10.1080/03634523.2016.1202994.

Bolkan, S. and Griffin, D.J. (2018): 'Catch and hold: instructional interventions and their differential impact on student interest, attention, and autonomous motivation', *Communication Education*, 67(3), pp. 269–286. Available at: https://doi.org/10.1080/03634523.2018.1465193.

Bolkan, S., Griffin, D.J. and Goodboy, A.K. (2018): 'Humor in the classroom: the effects of integrated humor on student learning', *Communication Education*, 67(2), pp. 144–164. Available at: https://doi.org/10.1080/03634523.2017.1413199.

Bradley, M.M. and Lang, P.J. (1994): 'Measuring emotion: The self-assessment manikin and the semantic differential', *Journal of Behavior Therapy and Experimental Psychiatry*, 25(1), pp. 49–59. Available at: https://doi.org/10.1016/0005-7916(94)90063-9.

Carroll, J.B. (1963): 'A Model of School Learning', *Teachers College Record: The Voice of Scholarship in Education*, 64(8), pp. 1–9. Available at: https://doi.org/10.1177/016146816306400801.

Carver, C.S. and Harmon-Jones, E. (2009): 'Anger is an approach-related affect: Evidence and implications.', *Psychological Bulletin*, 135(2), pp. 183–204. Available at: https://doi.org/10.1037/a0013965.

Carver, C.S. and Scheier, M.F. (2014): 'The experience of emotions during goal pursuit', *International handbook of emotions in education*. New York, NY, US: Routledge/Taylor & Francis Group (Educational psychology handbook series), pp. 56–72.

Çeken, B. and Taşkın, N. (2022): 'Multimedia learning principles in different learning environments: a systematic review', *Smart Learning Environments*, 9(1), pp. 1–22. Available at: https://doi.org/10.1186/s40561-022-00200-2.

Chiu, M.-H. et al. (2014): 'The role of facial microexpression state (FMES) change in the process of conceptual conflict', *British Journal of Educational Technology*, 45(3), pp. 471–486. Available at: https://doi.org/10.1111/bjet.12126.

Christensen, D.A., Brewer, B.W. and Hutchinson, J.C. (2018): *The Application of Video Data in a Multimodal Approach to Sport Psychology Research*. 1 Oliver's Yard, 55 City Road, London EC1Y 1SP United Kingdom: SAGE Publications Ltd. Available at: https://doi.org/10.4135/9781526438324.

Christenson, S., Reschly, A.L. and Wylie, C. (eds.) (2012): *Handbook of research on student engagement*. New York: Springer.

Chrysafiadi, K. and Virvou, M. (2013): 'Student modeling approaches: A literature review for the last decade', *Expert Systems with Applications*, 40(11), pp. 4715–4729. Available at: https://doi.org/10.1016/j.eswa.2013.02.007.

Condit, C.M. (2006): 'Communication as Relationality', in G. Shepherd, J. St. John, and T. Striphas, *Communication as…: Perspectives on Theory*. 2455 Teller Road, Thousand Oaks California 91320 United States: SAGE Publications, Inc., pp. 3–12. Available at: https://doi.org/10.4135/9781483329055.n1.

Corvinus University of Budapest (2020): *Research Ethics Committee - documents*. Available at: https://www.uni-corvinus.hu/main-page/about-the-university/senate-university-committees/kutatasetikai-bizottsag/related-documents/?lang=en (Accessed: 31 January 2026).

Craig, R.T. (1999): 'Communication Theory as a Field', *Communication Theory*, 9(2), pp. 119–161. Available at: https://doi.org/10.1111/j.1468-2885.1999.tb00355.x.

Csépe, V., Győri, Miklós. and Ragó, A. (2007): *Általános pszichológia*. Budapest: Osiris.

Csikszentmihalyi M. (1975): *Beyond boredom and anxiety*. Available at: https://openlibrary.org/works/OL506558W/Beyond_boredom_and_anxiety (Accessed: 29 March 2025).

Csikszentmihalyi (1990): 'Flow: The Psychology of Optimal Experience', *Journal of Leisure Research*, 24(1). Available at: https://www.tandfonline.com/doi/abs/10.1080/00222216.1992.11969876 (Accessed: 29 March 2025).

Danckert, J. et al. (2018): 'Boredom: What is it good for?', *The Function of Emotions: When and Why Emotions Help Us*, pp. 93–119. Available at: https://doi.org/10.1007/978-3-319-77619-4_6.

Daschmann, E.C., Goetz, T. and Stupnisky, R.H. (2014): 'Exploring the antecedents of boredom: Do teachers know why students are bored?', *Teaching and Teacher Education*, 39, pp. 22–30. Available at: https://doi.org/10.1016/j.tate.2013.11.009.

Davidson, R.J. et al. (1994): 'How are emotions distinguished from moods, temperament, and other related affective constructs?', *The nature of emotion: Fundamental questions*. New York, NY, US: Oxford University Press (Series in affective science), pp. 49–96.

Delahunty, J. and Crawford, N. (2024): 'Belonging, wellbeing and engagement: perspectives of regional, rural and remote students', in C. Stone and S. O'Shea (eds.) *Research Handbook on Student Engagement in Higher Education*. Edward Elgar Publishing, pp. 133–148. Available at: https://doi.org/10.4337/9781035314294.00020.

Dewan, M.A.A., Murshed, M. and Lin, F. (2019): 'Engagement detection in online learning: a review', *Smart Learning Environments*, 6(1), p. 1. Available at: https://doi.org/10.1186/s40561-018-0080-z.

Dewey, J. (1944): *Democracy and education : an introduction to the philosophy of education | WorldCat.org*. Available at: https://search.worldcat.org/title/democracy-and-education-an-introduction-to-the-philosophy-of-education/oclc/10795997/viewport (Accessed: 19 March 2025).

Dey, A. and Gottlieb, J. (2019): 'Attention, information-seeking, and active sampling: Empirical evidence and applications for learning.', *The Cambridge handbook of motivation and learning.* New York, NY, US: Cambridge University Press (Cambridge handbooks in psychology.), pp. 183–208. Available at: https://doi.org/10.1017/9781316823279.010.

Dixson, M.D. et al. (2017a): 'Nonverbal immediacy behaviors and online student engagement: bringing past instructional research into the present virtual classroom', *Communication Education*, 66(1), pp. 37–53. Available at: https://doi.org/10.1080/03634523.2016.1209222.

Dixson, M.D. et al. (2017b): 'Nonverbal immediacy behaviors and online student engagement: bringing past instructional research into the present virtual classroom', *Communication Education*, 66(1), pp. 37–53. Available at: https://doi.org/10.1080/03634523.2016.1209222.

D'Mello, S. et al. (2014): 'Confusion can be beneficial for learning', *Learning and Instruction*, 29, pp. 153–170. Available at: https://doi.org/10.1016/j.learninstruc.2012.05.003.

D'Mello, S., Dieterle, E. and Duckworth, A. (2017a): 'Advanced, Analytic, Automated (AAA) Measurement of Engagement During Learning', *Educational Psychologist*, 52(2), pp. 104–123. Available at: https://doi.org/10.1080/00461520.2017.1281747.

D'Mello, S., Dieterle, E. and Duckworth, A. (2017b): 'Advanced, Analytic, Automated (AAA) Measurement of Engagement During Learning', *Educational Psychologist*, 52(2), pp. 104–123. Available at: https://doi.org/10.1080/00461520.2017.1281747.

D'Mello, S. and Graesser, A. (2011): 'The half-life of cognitive-affective states during complex learning', *Cognition and Emotion*, 25(7), pp. 1299–1308. Available at: https://doi.org/10.1080/02699931.2011.613668.

D'Mello, S. and Kory, J. (2012): 'Consistent but modest: a meta-analysis on unimodal and multimodal affect detection accuracies from 30 studies', *Proceedings of the 14th ACM international conference on Multimodal interaction*. New York, NY, USA: Association for Computing Machinery (ICMI '12), pp. 31–38. Available at: https://doi.org/10.1145/2388676.2388686.

Dubbaka, A. and Gopalan, A. (2020): 'Detecting Learner Engagement in MOOCs using Automatic Facial Expression Recognition', *2020 IEEE Global Engineering Education Conference (EDUCON). 2020 IEEE Global Engineering Education Conference (EDUCON)*, pp. 447–456. Available at: https://doi.org/10.1109/EDUCON45650.2020.9125149.

Eastwood, J.D. et al. (2012): 'The Unengaged Mind: Defining Boredom in Terms of Attention', *Perspectives on Psychological Science*, 7(5), pp. 482–495. Available at: https://doi.org/10.1177/1745691612456044.

Ekman, P. (1970): 'Universal facial expressions of emotion', *California Mental Health Research Digest*, 8(4), pp. 151–158.

Ekman, P. (1992): ‘An argument for basic emotions’, *Cognition and Emotion*, 6(3–4), pp. 169–200. Available at: https://doi.org/10.1080/02699939208411068.

Ekman, P. (2003): *Emotions revealed: Recognizing faces and feelings to improve communication and emotional life*. New York, NY, US: Times Books/Henry Holt and Co (Emotions revealed: Recognizing faces and feelings to improve communication and emotional life), pp. xvii, 267.

Ekman, P. and Cordaro, D. (2011): ‘What is Meant by Calling Emotions Basic’, *Emotion Review*, 3(4), pp. 364–370. Available at: https://doi.org/10.1177/1754073911410740.

Ekman, P. and Friesen, W.V. (1975): *Unmasking the face: A guide to recognizing emotions from facial clues*. Oxford, England: Prentice-Hall (Unmasking the face: A guide to recognizing emotions from facial clues), pp. xii, 212.

Ekman, P. and Friesen, W.V. (1978): *Facial Action Coding System*. Consulting Psychologists Press (1. k.). Available at: https://books.google.hu/books?id=08l6wgEACAAJ.

Elpidorou, A. (2018): ‘The bored mind is a guiding mind: toward a regulatory theory of boredom’, *Phenomenology and the Cognitive Sciences*, 17(3), pp. 455–484. Available at: https://doi.org/10.1007/s11097-017-9515-1.

Elpidorou, A. (2021): ‘Is boredom one or many? A functional solution to the problem of heterogeneity’, *Mind & Language*, 36(3), pp. 491–511. Available at: https://doi.org/10.1111/mila.12282.

Elpidorou, A. (2024): ‘The Nature and Value of Boredom’, *The Routledge International Handbook of Boredom*. Routledge.

Fahlman, S.A. et al. (2013): ‘Development and Validation of the Multidimensional State Boredom Scale’, *Assessment*, 20(1), pp. 68–85. Available at: https://doi.org/10.1177/1073191111421303.

Farmer, R. and Sundberg, N.D. (1986): ‘Boredom Proneness--The Development and Correlates of a New Scale’, *Journal of Personality Assessment*, 50(1), pp. 4–17. Available at: https://doi.org/10.1207/s15327752jpa5001_2.

Farris, K.L., Houser, M.L. and Hosek, A.M. (2017): ‘Historical Roots and Trajectories of Instructional Communication’, *Handbook of Instructional Communication*. 2nd edn. Routledge.

Finkielsztein, M. (2020): ‘Class-related academic boredom among university students: a qualitative research on boredom coping strategies’, *Journal of Further and Higher Education*, 44(8), pp. 1098–1113. Available at: https://doi.org/10.1080/0309877X.2019.1658729.

Finkielsztein, M. (2024): ‘The Essence of Boredom: The Definition of Situational Boredom’, *Journal of Boredom Studies* [Preprint], (2). Available at: https://boredomsociety.com/jbs/index.php/journal/article/view/12 (Accessed: 19 July 2025).

Fiorella, L. (2021): ‘The Embodiment Principle in Multimedia Learning’, in R.E. Mayer and L. Fiorella (eds.) *The Cambridge Handbook of Multimedia Learning*. 3rd edn. Cambridge University Press, pp. 286–295. Available at: https://doi.org/10.1017/9781108894333.030.

Fredricks, J.A. (2022a): ‘The Measurement of Student Engagement: Methodological Advances and Comparison of New Self-report Instruments’, in A.L. Reschly and S.L. Christenson (eds.) *Handbook of Research on Student Engagement*. Cham: Springer International Publishing, pp. 597–616. Available at: https://doi.org/10.1007/978-3-031-07853-8_29.

Fredricks, J.A. (2022b): ‘The Measurement of Student Engagement: Methodological Advances and Comparison of New Self-report Instruments’, in A.L. Reschly and S.L. Christenson (eds.) *Handbook of Research on Student Engagement*. Cham: Springer International Publishing, pp. 597–616. Available at: https://doi.org/10.1007/978-3-031-07853-8_29.

Fredricks, J.A., Blumenfeld, P.C. and Paris, A.H. (2004): 'School Engagement: Potential of the Concept, State of the Evidence', *Review of Educational Research*, 74(1), pp. 59–109. Available at: https://doi.org/10.3102/00346543074001059.

Fredrickson, B.L. (1998): 'What Good Are Positive Emotions?', *Review of General Psychology*, 2(3), pp. 300–319. Available at: https://doi.org/10.1037/1089-2680.2.3.300.

Frenzel, A.C. et al. (2018): 'Emotion transmission in the classroom revisited: A reciprocal effects model of teacher and student enjoyment', *Journal of Educational Psychology*, 110(5), pp. 628–639. Available at: https://doi.org/10.1037/edu0000228.

Fyfield, M., Henderson, M. and Phillips, M. (2022): 'Improving instructional video design: A systematic review', *Australasian Journal of Educational Technology*, pp. 150–178. Available at: https://doi.org/10.14742/ajet.7296.

Garrison, D.R. (2018): *The Community of Inquiry: Shared Metacognition*, *The Community of Inquiry*. Available at: https://www.thecommunityofinquiry.org/editorial16 (Accessed: 16 March 2025).

Garrison, D.R., Anderson, T. and Archer, W. (1999): 'Critical Inquiry in a Text-Based Environment: Computer Conferencing in Higher Education', *The Internet and Higher Education*, 2(2–3), pp. 87–105. Available at: https://doi.org/10.1016/S1096-7516(00)00016-6.

Ghamdi, A.A., Samarji, A. and Watt, A. (2016): 'Essential Considerations in Distance Education in KSA: Teacher Immediacy in a Virtual Teaching and Learning Environment', *International Journal of Information and Education Technology*, 6(1), pp. 17–22. Available at: https://doi.org/10.7763/IJIET.2016.V6.651.

Goetz, T. et al. (2014): 'Types of boredom: An experience sampling approach.', *Motivation and Emotion*, 38(3), pp. 401–419. Available at: https://doi.org/10.1007/s11031-013-9385-y.

Goetz, T. et al. (2023): 'Test boredom: Exploring a neglected emotion.', *Journal of Educational Psychology*, 115(7), pp. 911–931. Available at: https://doi.org/10.1037/edu0000807.

Goetz et al. (2024): 'Academic Boredom', *The Routledge International Handbook of Boredom*, pp. 225–249. Available at: https://doi.org/10.4324/9781003271536-18.

Goetz, T. et al. (2024): 'Academic Boredom', *The Routledge International Handbook of Boredom*. Routledge.

Goetz, T., Hall, N.C. and Krannich, M. (2019a): 'Boredom', *The Cambridge Handbook of Motivation and Learning*. Cambridge University Press, pp. 465–489. Available at: https://doi.org/10.1017/9781316823279.021.

Goetz, T., Hall, N.C. and Krannich, M. (2019b): 'Boredom', *The Cambridge handbook of motivation and learning*. New York, NY, US: Cambridge University Press (Cambridge handbooks in psychology), pp. 465–489. Available at: https://doi.org/10.1017/9781316823279.021.

van Gog, T. (2021): 'The Signaling (or Cueing) Principle in Multimedia Learning', in L. Fiorella and R.E. Mayer (eds.) *The Cambridge Handbook of Multimedia Learning*. 3rd edn. Cambridge: Cambridge University Press (Cambridge Handbooks in Psychology), pp. 221–230. Available at: https://doi.org/10.1017/9781108894333.022.

Goren, C.C., Sarty, M. and Wu, P.Y.K. (1975): 'Visual Following and Pattern Discrimination of Face-like Stimuli by Newborn Infants', *Pediatrics*, 56(4), pp. 544–549. Available at: https://doi.org/10.1542/peds.56.4.544.

Gorham, J. (1988): 'The relationship between verbal teacher immediacy behaviors and student learning', *Communication Education*, 37(1), pp. 40–53. Available at: https://doi.org/10.1080/03634528809378702.

Grafsgaard, J.F. (2014): 'Multimodal analysis and modeling of nonverbal behaviors during tutoring'. *ICMI 2014 - Proceedings of the 2014 International Conference on Multimodal Interaction*, pp. 404–408. Available at: https://doi.org/10.1145/2663204.2667611.

Greene, B.A. (2015): 'Measuring Cognitive Engagement With Self-Report Scales: Reflections From Over 20 Years of Research', *Educational Psychologist*, 50(1), pp. 14–30. Available at: https://doi.org/10.1080/00461520.2014.989230.

Greene, J.C., Caracelli, V.J. and Graham, W.F. (1989): 'Toward a Conceptual Framework for Mixed-Method Evaluation Designs', *Educational Evaluation and Policy Analysis*, 11(3), pp. 255–274. Available at: https://doi.org/10.3102/01623737011003255.

Gudi, A. (2017): 'Affective Attitudes in FaceReader 7.1'. VicarVision.

Gyollai, A. et al. (2011): 'Psychometric properties of the Hungarian version of the original and the short form of the Positive and Negative Affect Schedule (PANAS)', *Neuropsychopharmacologia Hungarica : a Magyar Pszichofarmakológiai Egyesület lapja = official journal of the Hungarian Association of Psychopharmacology*, 13, pp. 73–9.

Han, J. and Patterson, I. (2007): 'An Analysis of the Influence That Leisure Experiences Have on a Person's Mood State, Health and Wellbeing', *Annals of Leisure Research*, 10(3–4), pp. 328–351. Available at: https://doi.org/10.1080/11745398.2007.9686770.

Harley, J.M. et al. (2019): 'I've got this: Fostering topic and technology-related emotional engagement and queer history knowledge with a mobile app', *Contemporary Educational Psychology*, 59. Available at: https://doi.org/10.1016/j.cedpsych.2019.101790.

Haxby, J.V., Hoffman, E.A. and Gobbini, M.I. (2000): 'The distributed human neural system for face perception', *Trends in Cognitive Sciences*, 4(6), pp. 223–233. Available at: https://doi.org/10.1016/S1364-6613(00)01482-0.

Heckhausen, H. (1977): 'Achievement motivation and its constructs: A cognitive model', *Motivation and Emotion*, 1(4), pp. 283–329. Available at: https://doi.org/10.1007/BF00992538.

Herrando, C. and Constantinides, E. (2021): 'Emotional Contagion: A Brief Overview and Future Directions', *Frontiers in Psychology*, 12, p. 712606. Available at: https://doi.org/10.3389/fpsyg.2021.712606.

Hetland, A. et al. (2019): 'The thrill of speedy descents: A pilot study on differences in facially expressed online emotions and retrospective measures of emotions during a downhill mountain-bike descent', *Frontiers in Psychology*, 10(APR). Available at: https://doi.org/10.3389/fpsyg.2019.00566.

Hidi, S. and Harackiewicz, J.M. (2000): 'Motivating the Academically Unmotivated: A Critical Issue for the 21st Century', *Review of Educational Research*, 70(2), pp. 151–179. Available at: https://doi.org/10.3102/00346543070002151.

Hirt, F.S. et al. (2018): 'A comparison of students' emotional self-reports with automated facial emotion recognition in a reading situation', *undefined* [Preprint]. Available at: https://www.semanticscholar.org/paper/A-comparison-of-students%27-emotional-self-reports-in-Hirt-Moser/4cb34f37a433916e0f49cc711888a123693e69b4 (Accessed: 7 May 2022).

Horan, S.M., Martin, M.M. and Weber, K. (2012): 'Understanding Emotional Response Theory: The Role of Instructor Power and Justice Messages', *Communication Quarterly*, 60(2), pp. 210–233. Available at: https://doi.org/10.1080/01463373.2012.669323.

Houser, M.L. and Hosek, A.M. (eds.) (2022): *Handbook of instructional communication: principles & practices of teaching rhetorical & relational perspectives*. Third edition. Dubuque, IA: Kendall Hunt.

Huggett, C. (2021): ‘The state of virtual training 2022’, *https://www.cindyhuggett.com/*, 7 0. Available at: https://www.cindyhuggett.com/blog/2022sovt/.

Hunter, J.A. and Eastwood, J.D. (2021): ‘Understanding the relation between boredom and academic performance in postsecondary students.’, *Journal of Educational Psychology*, 113(3), pp. 499–515. Available at: https://doi.org/10.1037/edu0000479.

Iphofen, R. (2011): *Ethical Decision-Making in Social Research*. London: Palgrave Macmillan UK. Available at: https://doi.org/10.1057/9780230233768.

Jang, E.-H. et al. (2015): ‘Analysis of physiological signals for recognition of boredom, pain, and surprise emotions’, *Journal of Physiological Anthropology*, 34(1), p. 25. Available at: https://doi.org/10.1186/s40101-015-0063-5.

Jang, H., Kim, E.J. and Reeve, J. (2016): ‘Why students become more engaged or more disengaged during the semester: A self-determination theory dual-process model’, *Learning and Instruction*, 43, pp. 27–38. Available at: https://doi.org/10.1016/j.learninstruc.2016.01.002.

Jewitt, C. (2014): ‘What is Multimodal Research?’ SAGE Publications Ltd. Available at: https://dx.doi.org/10.4135/9781473907669.

Jia, X. et al. (2020): ‘The role of social capital on proactive and reactive resilience of organizations post-disaster’, *International Journal of Disaster Risk Reduction*, 48, p. 101614. Available at: https://doi.org/10.1016/j.ijdrr.2020.101614.

Kahu, E.R. (2013): ‘Framing student engagement in higher education’, *Studies in Higher Education*, 38(5), pp. 758–773. Available at: https://doi.org/10.1080/03075079.2011.598505.

Kahu, E. (2024): ‘Foreword: developing a conceptual framework of student engagement’. Available at: https://www.elgaronline.com/edcollchap/book/9781035314294/front-8.xml (Accessed: 24 April 2025).

Kahu, E.R. and Nelson, K. (2018): ‘Student engagement in the educational interface: understanding the mechanisms of student success’, *Higher Education Research & Development*, 37(1), pp. 58–71. Available at: https://doi.org/10.1080/07294360.2017.1344197.

Kellas, J.K. (2008): ‘Narrative Theories: Making Sense of Interpersonal Communication’, in L. Baxter and D. Braithwaite, *Engaging Theories in Interpersonal Communication: Multiple Perspectives*. 2455 Teller Road, Thousand Oaks California 91320 United States: SAGE Publications, Inc., pp. 241–254. Available at: https://doi.org/10.4135/9781483329529.n18.

Keltner, D. (2016): *Understanding Multimodal Emotional Expressions: Recent Advances in Basic Emotion Theory | Emotion Researcher*. Available at: http://emotionresearcher.com/understanding-multimodal-emotional-expressions-recent-advances-in-basic-emotion-theory/ (Accessed: 28 June 2024).

Keltner, D. et al. (2019): ‘Emotional Expression: Advances in Basic Emotion Theory’, *Journal of Nonverbal Behavior*, 43(2), pp. 133–160. Available at: https://doi.org/10.1007/s10919-019-00293-3.

Keltner, D. and Cordaro, D.T. (2017): *Understanding Multimodal Emotional Expressions*. Oxford University Press. Available at: https://doi.org/10.1093/acprof:oso/9780190613501.003.0004.

Kift, S. (2024): ‘The quest for the holy grail of student engagement: curriculum as the engagement integrator’, in C. Stone and S. O’Shea (eds.) *Research Handbook on Student Engagement in Higher Education*. Edward Elgar Publishing, pp. 7–27. Available at: https://doi.org/10.4337/9781035314294.00011.

Krathwohl, D.R., Bloom, B.S. and Masia, B.B. (1964): *Taxonomy Of Educational Objectives, Handbook II: Affective Domain*.

Kromka, S.M. and Goodboy, A.K. (2019): 'Classroom storytelling: using instructor narratives to increase student recall, affect, and attention', *Communication Education*, 68(1), pp. 20–43. Available at: https://doi.org/10.1080/03634523.2018.1529330.

La Torre, S. and Désiron, J.C. (2024): 'From Research to Practice: Are Multimedia Principles Present in Instructional Videos Used by Teachers in Science and History?', *Technology, Knowledge and Learning*, 29(4), pp. 1993–2016. Available at: https://doi.org/10.1007/s10758-024-09753-2.

Landmann, E. (2023): 'I can see how you feel—Methodological considerations and handling of Noldus's FaceReader software for emotion measurement', *Technological Forecasting and Social Change*, 197, p. 122889. Available at: https://doi.org/10.1016/j.techfore.2023.122889.

Langton, S.R.H., Watt, R.J. and Bruce, V. (2000): 'Do the eyes have it? Cues to the direction of social attention', *Trends in Cognitive Sciences*, 4(2), pp. 50–59. Available at: https://doi.org/10.1016/S1364-6613(99)01436-9.

Lazarus, R.S. (1991): *Emotion and adaptation*. New York, NY, US: Oxford University Press (Emotion and adaptation), pp. xiii, 557.

Lewinski, P. (2015a): 'Automated facial coding software outperforms people in recognizing neutral faces as neutral from standardized datasets', *Frontiers in Psychology*, 6. Available at: https://doi.org/10.3389/fpsyg.2015.01386.

Lewinski, P. (2015b): 'Don't look blank, happy, or sad: Patterns of facial expressions of speakers in banks' YouTube videos predict video's popularity over time.', *Journal of Neuroscience, Psychology, and Economics*, 8(4), pp. 241–249. Available at: https://doi.org/10.1037/npe0000046.

Lewinski, P., Den Uyl, T.M. and Butler, C. (2014): 'Automated facial coding: Validation of basic emotions and FACS AUs in FaceReader.', *Journal of Neuroscience, Psychology, and Economics*, 7(4), pp. 227–236. Available at: https://doi.org/10.1037/npe0000028.

Lewinski, P., Fransen, M.L. and Tan, E.S.H. (2014): 'Predicting advertising effectiveness by facial expressions in response to amusing persuasive stimuli.', *Journal of Neuroscience, Psychology, and Economics*, 7(1), pp. 1–14. Available at: https://doi.org/10.1037/npe0000012.

Li, J. et al. (2016): 'Multimodal human attention detection for reading from facial expression, eye gaze, and mouse dynamics', *ACM SIGAPP Applied Computing Review*, 16(3), pp. 37–49. Available at: https://doi.org/10.1145/3015297.3015301.

Liu, X.-X. et al. (2021): 'Perceived teacher support and creative self-efficacy: The mediating roles of autonomous motivation and achievement emotions in Chinese junior high school students', *Thinking Skills and Creativity*, 39. Available at: https://doi.org/10.1016/j.tsc.2020.100752.

Loijens, L. and Krips, O.E. (2025): *FaceReader methodology note*. White paper. Wageningen: Noldus Information Technology. Available at: https://info.noldus.com/hubfs/resources/noldus-white-paper-facereader-methodology.pdf.

Martarelli, C.S. and Jost, N.S. (2024): 'Methodological approaches to boredom and its measurement', *The Routledge International Handbook of Boredom*. Routledge.

Matthews, K.E., Lodge, J.M. and Johnstone, M. (2024): 'The relational turn in student engagement', in C. Stone and S. O'Shea (eds.) *Research Handbook on Student Engagement in Higher Education*. Edward Elgar Publishing, pp. 28–38. Available at: https://doi.org/10.4337/9781035314294.00012.

Mayer, R.E. (2021): ‘Cognitive Theory of Multimedia Learning’, in R.E. Mayer and L. Fiorella (eds.) *The Cambridge Handbook of Multimedia Learning*. 3rd edn. Cambridge University Press, pp. 57–72. Available at: https://doi.org/10.1017/9781108894333.008.

Mayer, R.E. and Fiorella, L. (eds.) (2021): *The Cambridge Handbook of Multimedia Learning*. 3rd edn. Cambridge University Press. Available at: https://doi.org/10.1017/9781108894333.

Mazer, J.P. (2012): ‘Development and Validation of the Student Interest and Engagement Scales’, *Communication Methods and Measures*, 6(2), pp. 99–125. Available at: https://doi.org/10.1080/19312458.2012.679244.

Mazer, J.P. (2013a): ‘Student Emotional and Cognitive Interest as Mediators of Teacher Communication Behaviors and Student Engagement: An Examination of Direct and Interaction Effects’, *Communication Education*, 62(3), pp. 253–277. Available at: https://doi.org/10.1080/03634523.2013.777752.

Mazer, J.P. (2013b): ‘Student Emotional and Cognitive Interest as Mediators of Teacher Communication Behaviors and Student Engagement: An Examination of Direct and Interaction Effects’, *Communication Education*, 62(3), pp. 253–277. Available at: https://doi.org/10.1080/03634523.2013.777752.

Mazer, J.P. (2017): ‘Associations among classroom emotional processes, student interest, and engagement: a convergent validity test’, *Communication Education*, 66(3), pp. 350–360. Available at: https://doi.org/10.1080/03634523.2016.1265134.

Mazer, J.P. (2022): ‘Instructional message variables’, in M.L. Houser and A.M. Hosek (eds.) *Handbook of Instructional Communication: Principles and Practices of Teaching Rhetorical and Relational Perspectives*. 3rd edn. Kendall Hunt, pp. 29-54.

McDaniel, B. et al. (2007): ‘Facial Features for Affective State Detection in Learning Environments’, *Proceedings of the Annual Meeting of the Cognitive Science Society*, 29(29). Available at: https://escholarship.org/uc/item/9w00945d (Accessed: 28 April 2023).

Mehrabian, A. (1981): *Silent Messages*. Available at: https://www.academia.edu/23744443/Albert_Mehrabian_Silent_Messages (Accessed: 13 February 2025).

Mercer-Mapstone, L. et al. (2024): *Recommendations for equitable student support during disruptions to the higher education sector: Lessons from COVID-19*. CQUniversity, p. 0 Bytes. Available at: https://doi.org/10.25946/26018914.

Merrifield, C. and Danckert, J. (2014): ‘Characterizing the psychophysiological signature of boredom’, *Experimental Brain Research*, 232(2), pp. 481–491. Available at: https://doi.org/10.1007/s00221-013-3755-2.

Mikulas, W.L. and Vodanovich, S.J. (1993): ‘The essence of boredom’, *The Psychological Record*, 43(1), pp. 3–12.

Mills, C. et al. (2018): ‘Is an off-task mind a freely-moving mind? Examining the relationship between different dimensions of thought’, *Consciousness and Cognition*, 58, pp. 20–33. Available at: https://doi.org/10.1016/j.concog.2017.10.003.

Mottet, T., Frymier, A.B. and Beebe, S.A. (2006): ‘Theorizing About Instructional Communication’, *Handbook of instructional communication: Rhetorical and relational perspectives*. Routledge, pp. 255–282.

Mugon, J., Danckert, J. and Eastwood, J. (2019): ‘The costs and benefits of boredom in the classroom’, *The Cambridge Handbook of Motivation and Learning*, pp. 490–514. Available at:

https://www.scopus.com/inward/record.uri?eid=2-s2.0-85097803955&partnerID=40&md5=b115178ce9106002388cacb4fdaa2fdc.

Neuroskeptic (2020): *How Has The Pandemic Affected Mental Health Research?*, *Discover Magazine*. Available at: https://www.discovermagazine.com/the-sciences/how-has-the-pandemic-affected-mental-health-research (Accessed: 20 March 2025).

Noroozi, O. et al. (2020): 'Multimodal data indicators for capturing cognitive, motivational, and emotional learning processes: A systematic literature review', *Education and Information Technologies*, 25(6), pp. 5499–5547. Available at: https://doi.org/10.1007/s10639-020-10229-w.

O'Connor, C. and Joffe, H. (2020): 'Intercoder Reliability in Qualitative Research: Debates and Practical Guidelines', *International Journal of Qualitative Methods*, 19, p. 1609406919899220. Available at: https://doi.org/10.1177/1609406919899220.

OECD (2019): *OECD Learning Compass 2030: A Series of Concept Notes*. Paris: OECD. Available at: https://www.oecd.org/en/data/tools/oecd-learning-compass-2030.html (Accessed: 30 August 2026).

Ohsuga, M., Shimono, F. and Genno, H. (2001): 'Assessment of phasic work stress using autonomic indices', *International Journal of Psychophysiology*, 40(3), pp. 211–220. Available at: https://doi.org/10.1016/S0167-8760(00)00189-6.

O'Shea, S. and Delahunty, J. (2018): '*Getting through the day and still having a smile on my face* ! How do students define success in the university learning environment?', *Higher Education Research & Development*, 37(5), pp. 1062–1075. Available at: https://doi.org/10.1080/07294360.2018.1463973.

Parker, P.C. et al. (2018): 'An attribution-based motivation treatment for low control students who are bored in online learning environments.', *Motivation Science*, 4(2), pp. 177–184. Available at: https://doi.org/10.1037/mot0000081.

Pattyn, N. et al. (2008): 'Psychophysiological investigation of vigilance decrement: Boredom or cognitive fatigue?', *Physiology & Behavior*, 93(1–2), pp. 369–378. Available at: https://doi.org/10.1016/j.physbeh.2007.09.016.

Pekrun, R. (2006): 'The Control-Value Theory of Achievement Emotions: Assumptions, Corollaries, and Implications for Educational Research and Practice', *Educational Psychology Review*, 18(4), pp. 315–341. Available at: https://doi.org/10.1007/s10648-006-9029-9.

Pekrun, R. et al. (2010): 'Boredom in Achievement Settings: Exploring Control-Value Antecedents and Performance Outcomes of a Neglected Emotion', *Journal of Educational Psychology*, 102(3), pp. 531–549. Available at: https://doi.org/10.1037/a0019243.

Pekrun, R. et al. (2014): 'Boredom and Academic Achievement: Testing a Model of Reciprocal Causation', *Journal of Educational Psychology*, 106(3), pp. 696–710. Available at: https://doi.org/10.1037/a0036006.

Pekrun, R. (2019): 'Expectancy–Value Theory of Anxiety: Overview and Implications', in T. Sosnowski and K. Wrzesniewski, *Anxiety: Recent Developments in Cognitive, Psychophysiological, and Health Research*. 1st edn. Edited by D. G. Forgays. Taylor & Francis, pp. 23–42. Available at: https://doi.org/10.4324/9781315825199-3.

Pekrun, R., Marsh, H.W., Elliot, A.J., et al. (2023): 'A three-dimensional taxonomy of achievement emotions.', *Journal of Personality and Social Psychology*, 124(1), pp. 145–178. Available at: https://doi.org/10.1037/pspp0000448.

Pekrun, R., Marsh, H.W., Suessenbach, F., et al. (2023): 'School grades and students' emotions: Longitudinal models of within-person reciprocal effects', *Learning and Instruction*, 83, p. 101626. Available at: https://doi.org/10.1016/j.learninstruc.2022.101626.

Pekrun, R. (2024): 'Control-Value Theory: From Achievement Emotion to a General Theory of Human Emotions', *Educational Psychology Review*, 36(3), p. 83. Available at: https://doi.org/10.1007/s10648-024-09909-7.

Pekrun, R. and Goetz, T. (2024): 'Boredom: A Control-Value Theory approach', *The Routledge International Handbook of Boredom*. Routledge.

Pekrun, R. and Linnenbrink-Garcia, L. (2022): 'Academic Emotions and Student Engagement', in A.L. Reschly and S.L. Christenson (eds.) *Handbook of Research on Student Engagement*. Cham: Springer International Publishing, pp. 109–132. Available at: https://doi.org/10.1007/978-3-031-07853-8_6.

Phillips, D.C. (1995): 'The Good, the Bad, and the Ugly: The Many Faces of Constructivism', *Educational Researcher*, 24(7), pp. 5–12. Available at: https://doi.org/10.3102/0013189X024007005.

Plass, J.L. and Hovey, C. (2021): 'The Emotional Design Principle in Multimedia Learning', in R.E. Mayer and L. Fiorella (eds.) *The Cambridge Handbook of Multimedia Learning*. 3rd edn. Cambridge University Press, pp. 324–336. Available at: https://doi.org/10.1017/9781108894333.034.

Plutchik, R. (1980): 'A GENERAL PSYCHOEVOLUTIONARY THEORY OF EMOTION', *Theories of Emotion*. Elsevier, pp. 3–33. Available at: https://doi.org/10.1016/B978-0-12-558701-3.50007-7.

Poria, S. (2016): 'Fusing audio, visual and textual clues for sentiment analysis from multimodal content', *Neurocomputing*, 174, pp. 50–59.

Redmond, P. et al. (2018): 'An Online Engagement Framework for Higher Education', *Online Learning*, 22(1). Available at: https://doi.org/10.24059/olj.v22i1.1175.

Reschly, A.R. and Christenson, S.L. (2006): 'Promoting successful school completion: Development, Prevention, and Intervention', in G.G. Bear and K.M. Minke (eds.) *Children's Needs-III*. Bethesda, MD: National Association of School Psychologists, pp. 103–113.

Reschly, A.L. and Christenson, S.L. (2012): 'Jingle, Jangle, and Conceptual Haziness: Evolution and Future Directions of the Engagement Construct', in S.L. Christenson, A.L. Reschly, and C. Wylie (eds.) *Handbook of Research on Student Engagement*. Boston, MA: Springer US, pp. 3–19. Available at: https://doi.org/10.1007/978-1-4614-2018-7_1.

Reschly, A.L. and Christenson, S.L. (2022): 'Jingle-Jangle Revisited: History and Further Evolution of the Student Engagement Construct', in A.L. Reschly and S.L. Christenson (eds.) *Handbook of Research on Student Engagement*. Cham: Springer International Publishing, pp. 3–24. Available at: https://doi.org/10.1007/978-3-031-07853-8_1.

Richmond, V.P., Gorham, J.S. and Mccroskey, J.C. (1987): 'The Relationship Between Selected Immediacy Behaviors and Cognitive Learning', *Annals of the International Communication Association*, 10(1), pp. 574–590. Available at: https://doi.org/10.1080/23808985.1987.11678663.

Russell, J.A. (1980): 'A circumplex model of affect.', *Journal of Personality and Social Psychology*, 39(6), pp. 1161–1178. Available at: https://doi.org/10.1037/h0077714.

Ryan, R.M. and Deci, E.L. (2000): 'Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being.', *American Psychologist*, 55(1), pp. 68–78. Available at: https://doi.org/10.1037/0003-066X.55.1.68.

Salmela-Aro, K., Tang, X. and Upadyaya, K. (2022): 'Study Demands-Resources Model of Student Engagement and Burnout', in A.L. Reschly and S.L. Christenson (eds.) *Handbook of Research on Student Engagement*. Cham: Springer International Publishing, pp. 77–93. Available at: https://doi.org/10.1007/978-3-031-07853-8_4.

Sánchez-Rosas, J. and Esquivel, S. (2016): 'Instructional Teaching Quality, Task Value, Self-Efficacy, and Boredom: A Model of Attention in Class', *Revista de Psicología*, 25(2), pp. 1–20. Available at: https://www.redalyc.org/articulo.oa?id=26449350011 (Accessed: 9 June 2023).

Sass, J. and Vinczéné Fekete, L. (2022): 'Methods for Measuring Student Emotions during Online Learning', *PRACTICE AND THEORY IN SYSTEMS OF EDUCATION*, 17, pp. 95–113. Available at: https://m2.mtmt.hu/api/publication/33712852.

Sass, J., Vinczéné Fekete, L. and Bodnár, É. (2023): 'Boredom as a Challenge to Knowledge Acquisition: Mapping Boredom-Inducing Elements of a Video Presentation, Using Multimodal Measurements', in A. Lerro, D. Carlucci, and G. Schiuma (eds.) *Proceedings IFKAD: Managing Knowledge for Sustainability*. Potenza: University of Basilicata, pp. 2379–2393. Available at: https://m2.mtmt.hu/api/publication/34026526.

Scherer, K.R. (2005): 'What are emotions? And how can they be measured?', *Social Science Information*, 44(4), pp. 695–729. Available at: https://doi.org/10.1177/0539018405058216.

Scherer, K.R. and Moors, A. (2019): 'The Emotion Process: Event Appraisal and Component Differentiation', *Annual Review of Psychology*, 70(1), pp. 719–745. Available at: https://doi.org/10.1146/annurev-psych-122216-011854.

Schlechty, P.C. (2011): *Engaging Students: The Next Level of Working on the Work*. Hoboken: John Wiley & Sons, Incorporated.

Schrader, C., Kalyuga, S. and Plass, J.L. (2021): 'Motivation and Affect in Multimedia Learning', in R.E. Mayer and L. Fiorella (eds.) *The Cambridge Handbook of Multimedia Learning*. 3rd edn. Cambridge University Press, pp. 121–131. Available at: https://doi.org/10.1017/9781108894333.012.

Seneca (60): *On the Tranquility of the Mind*, *Goodreads*. Available at: https://www.goodreads.com/book/show/33597739-peace-of-mind (Accessed: 19 March 2025).

Sharp, J.G. et al. (2018): 'Academic boredom, approaches to learning and the final-year degree outcomes of undergraduate students', *Journal of Further and Higher Education*, 42(8), pp. 1055–1077. Available at: https://doi.org/10.1080/0309877X.2017.1349883.

Silvia, P.J. (2006): *Exploring the Psychology of Interest*. 1st edn. Oxford University PressNew York. Available at: https://doi.org/10.1093/acprof:oso/9780195158557.001.0001.

Sinatra, G.M., Heddy, B.C. and Lombardi, D. (2015): 'The Challenges of Defining and Measuring Student Engagement in Science', *Educational Psychologist*, 50(1), pp. 1–13. Available at: https://doi.org/10.1080/00461520.2014.1002924.

Skinner, E.A. and Raine, K.E. (2022): 'Unlocking the Positive Synergy Between Engagement and Motivation', in A.L. Reschly and S.L. Christenson (eds.) *Handbook of Research on Student Engagement*. Cham: Springer International Publishing, pp. 25–56. Available at: https://doi.org/10.1007/978-3-031-07853-8_2.

Smith, R.A. and Tinto, V. (2022): 'Unraveling Student Engagement: Exploring its Relational and Longitudinal Character', *Journal of College Student Retention: Research, Theory & Practice* [Preprint]. Available at: https://doi.org/10.1177/15210251221098172.

Sung, E. and Mayer, R.E. (2012): 'Five facets of social presence in online distance education', *Computers in Human Behavior*, 28(5), pp. 1738–1747. Available at: https://doi.org/10.1016/j.chb.2012.04.014.

Sweller, J. (1988): 'Cognitive Load During Problem Solving: Effects on Learning', *Cognitive Science*, 12(2), pp. 257–285. Available at: https://doi.org/10.1207/s15516709cog1202_4.

Talen, L. and Den Uyl, T.E. (2022): 'Complex Website Tasks Increase the Expression Anger Measured with FaceReader Online', *International Journal of Human–Computer Interaction*, 38(3), pp. 282–288. Available at: https://doi.org/10.1080/10447318.2021.1938390.

Tam, K.Y.Y. et al. (2021): 'Attention Drifting In and Out: The Boredom Feedback Model', *Personality and Social Psychology Review*, 25(3), pp. 251–272. Available at: https://doi.org/10.1177/10888683211010297.

Tan Ti-Shen, E. and Sen, A. (2019): 'EFFECTIVE USE OF INTERACTIVE VIDEOS AS AN EDUCATIONAL TOOL IN HIGHER EDUCATION – AN ASIAN CONTEXT'. *11th International Conference on Education and New Learning Technologies*, Palma, Spain, pp. 10379–10385. Available at: https://doi.org/10.21125/edulearn.2019.2610.

Terzis, V., Moridis, C.N. and Economides, A.A. (2012): 'The effect of emotional feedback on behavioral intention to use computer based assessment', *Computers & Education*, 59(2), pp. 710–721. Available at: https://doi.org/10.1016/j.compedu.2012.03.003.

Terzis, V., Moridis, C.N. and Economides, A.A. (2013): 'Measuring instant emotions based on facial expressions during computer-based assessment', *Personal and Ubiquitous Computing*, 17(1), pp. 43–52. Available at: https://doi.org/10.1007/s00779-011-0477-y.

Tilburg, W.A.P. van et al. (2024): 'Boredom Proneness', *The Routledge International Handbook of Boredom*. Routledge.

Titsworth, S. et al. (2013): 'The Bright Side of Emotion in the Classroom: Do Teachers' Behaviors Predict Students' Enjoyment, Hope, and Pride?', *Communication Education*, 62(2), pp. 191–209. Available at: https://doi.org/10.1080/03634523.2013.763997.

Trowler, V. (2015): 'Negotiating Contestations and "Chaotic Conceptions": Engaging "Non-Traditional" Students in Higher Education', *Higher Education Quarterly*, 69(3), pp. 295–310. Available at: https://doi.org/10.1111/hequ.12071.

Trowler, V. et al. (2022): 'Pathways to student engagement: beyond triggers and mechanisms at the engagement interface', *Higher Education*, 84(4), pp. 761–777. Available at: https://doi.org/10.1007/s10734-021-00798-1.

Tze, V., Parker, P. and Daniels, L. (2024): 'Boredom Intervention: Recommendations, experiences, and preliminary evidence', *The Routledge International Handbook of Boredom*. Routledge.

Van Tilburg, W.A.P. and Igou, E.R. (2012): 'On boredom: Lack of challenge and meaning as distinct boredom experiences', *Motivation and Emotion*, 36(2), pp. 181–194. Available at: https://doi.org/10.1007/s11031-011-9234-9.

Vaughan, N. et al. (2023): *Principles of Blended Learning: Shared Metacognition and Communities of Inquiry*. Available at: https://doi.org/10.15215/aupress/9781771993920.01.

Vogel-Walcutt, J.J. et al. (2012): 'The Definition, Assessment, and Mitigation of State Boredom Within Educational Settings: A Comprehensive Review', *Educational Psychology Review*, 24(1), pp. 89–111. Available at: https://doi.org/10.1007/s10648-011-9182-7.

Waldbuesser, C. and Titsworth, S. (2022): 'Communication Perspectives on Emotions in the Instructional Context', in M.L. Houser and A.M. Hosek (eds.) *Handbook of instructional*

communication: principles & practices of teaching rhetorical & relational perspectives. Third edition. Dubuque, IA: Kendall Hunt.

Wallace, J.C., Vodanovich, S.J. and Restino, B.M. (2003): 'Predicting cognitive failures from boredom proneness and daytime sleepiness scores: an investigation within military and undergraduate samples', *Personality and Individual Differences*, 34(4), pp. 635–644. Available at: https://doi.org/10.1016/S0191-8869(02)00050-8.

Wang, M.-T., Degol, J.L. and Henry, D.A. (2019): 'An integrative development-in-sociocultural-context model for children's engagement in learning.', *American Psychologist*, 74(9), pp. 1086–1102. Available at: https://doi.org/10.1037/amp0000522.

Wanzer, M.B. and Frymier, A.B. (1999): 'The relationship between student perceptions of instructor humor and students' reports of learning', *Communication Education*, 48(1), pp. 48–62. Available at: https://doi.org/10.1080/03634529909379152.

Watson, D., Clark, L.A. and Tellegen, A. (1988): 'Development and validation of brief measures of positive and negative affect: The PANAS scales', *Journal of Personality and Social Psychology*, 54(6), pp. 1063–1070. Available at: https://doi.org/10.1037/0022-3514.54.6.1063.

Watson, D. and Tellegen, A. (1985a): 'Toward a consensual structure of mood.', *Psychological Bulletin*, 98(2), pp. 219–235. Available at: https://doi.org/10.1037/0033-2909.98.2.219.

Watson, D. and Tellegen, A. (1985b): 'Toward a consensual structure of mood.', *Psychological Bulletin*, 98(2), pp. 219–235. Available at: https://doi.org/10.1037/0033-2909.98.2.219.

Weibel, D. and Wissmath, B. (2024): 'Flow As An Experience Beyond Boredom', *The Routledge International Handbook of Boredom*. Routledge.

Weiner, B. (1985): 'An attributional theory of achievement motivation and emotion', *Psychological Review*, 92(4), pp. 548–573. Available at: https://doi.org/10.1037/0033-295X.92.4.548.

Westgate, E.C. and Steidle, B. (2020): 'Lost by definition: Why boredom matters for psychology and society', *Social and Personality Psychology Compass*, 14(11). Available at: https://doi.org/10.1111/spc3.12562.

Westgate, E.C. and Wilson, T.D. (2018): 'Boring thoughts and bored minds: The MAC model of boredom and cognitive engagement.', *Psychological Review*, 125(5), pp. 689–713. Available at: https://doi.org/10.1037/rev0000097.

Whitehill, J. et al. (2014): 'The Faces of Engagement: Automatic Recognition of Student Engagementfrom Facial Expressions', *IEEE Transactions on Affective Computing*, 5(1), pp. 86–98. Available at: https://doi.org/10.1109/TAFFC.2014.2316163.

Williams, K.H., Childers, C. and Kemp, E. (2013): 'Stimulating and Enhancing Student Learning Through Positive Emotions', *Journal of Teaching in Travel & Tourism*, 13(3), pp. 209–227. Available at: https://doi.org/10.1080/15313220.2013.813320.

Wolff, W., Radtke, V.C. and Martarelli, C.S. (2024): 'Same same but different: What is boredom actually?', *The Routledge International Handbook of Boredom*. Routledge.

Woolfitt, Z. (2015): 'The effective use of video in higher education'. Available at: https://www.academia.edu/22387499/The_effective_use_of_video_in_higher_education (Accessed: 7 May 2022).

Wu, C.-H., Huang, Y.-M. and Hwang, J.-P. (2016): 'Review of affective computing in education/learning: Trends and challenges', *British Journal of Educational Technology*, 47(6), pp. 1304–1323. Available at: https://doi.org/10.1111/bjet.12324.

Yadegaridehkordi, E. et al. (2019): ‘Affective computing in education: A systematic review and future research’, *Computers & Education*, 142, p. 103649. Available at: https://doi.org/10.1016/j.compedu.2019.103649.

Zhong, B. et al. (2017): ‘Emotion recognition with facial expressions and physiological signals’, *2017 IEEE Symposium Series on Computational Intelligence (SSCI)*. *2017 IEEE Symposium Series on Computational Intelligence (SSCI)*, Honolulu, HI: IEEE, pp. 1–8. Available at: https://doi.org/10.1109/SSCI.2017.8285365.

13 AI DECLARATION

I acknowledge the use of AI to facilitate certain aspects of the literature review, specifically to generate the table summarising the main features of the studies and to improve the overall language quality of the manuscript, which significantly contributed to the efficiency and clarity of our analysis.

The adequacy and correctness of the thesis's academic language were enhanced using the following tools: ChatGPT, DeepL Translate, DeepL Write, and Grammarly.

14 APPENDICES

Appendix A: **Differences between student motivation and engagement**

	Student motivation	Student engagement
Concept type	It is discussed by a range of theories and their sub-theories.	It is rather a framework, or model
Nature and scope	It is a key element of student engagement, but it is a part of its broader construct. Motivational concepts are the indicators of engagement	A broad, interdisciplinary, integrated meta-construct
Educational sectors affected by empirical research	Its research primarily focuses on secondary and higher education	Covers all levels of education, from early childhood education to tertiary programmes
Application areas	Narrower, more suitable for theory testing, and can be better integrated with existing motivational theories.	It is more effective in communicating with policymakers, educators and parents.
Their role in everyday learning practices	The 'intention', preceding engagement	The 'action', outcome behaviours of motivation

Source: Author's own compilation, based on Christenson et al. 2012, Martin 2007, Reschly & Christenson 2022 and Skinner & Raine 2022

Appendix B: Overview of the context, methods and findings of studies applying FaceReader (FR)

	Citation	Scimago ranking	Context	Aim	Methods used (data collection, procedure, statistical analysis methods)	Sample size	Conclusion
1.	Terzis et al., 2012	Q1	education; computer based assessment, introductory informatics course	To explore the effect of emotional feedback on behavioural intention to use computer based assessment.	Data collection: - students completed a voluntary computer-based assessment (CBA) with 45 multiple-choice questions - followed by two questionnaires evaluating emotional feedback and related constructs. Analysis: Model Testing: Partial least squares (PLS) analysis was employed to evaluate the structural and measurement models. Reliability and Validity: Internal consistency, convergent validity (factor loadings ≥ 0.7), and discriminant validity (AVE > 0.5) were ensured. Goodness of Fit: R^2 values assessed the variance explained by predictors Bootstrap Analysis: Hypotheses were tested using 1,000 resamples to determine statistical significance.	134	Emotional Feedback is a significant determinant for Behavioral Intention to Use a CBA, Perceived Playfulness, Perceived Usefulness, Perceived Ease of Use, Content and Facilitating Conditions educators in order to enhance learner's intentions to use CBA systems, they should provide an emotional feedback during the CBA with careful and relevant design and content.
2.	Terzis et al., 2013	Q1	Education: task completion (multiple choice questions) 45 minutes, 45 questions	a) examination of FaceReader's efficiency in measuring students' instant emotions b) provide empirical data concerning students' instant emotions	Data collection: - 2 human coders behind the curtains & FR Analysis: - examined the frequency, not the length and extent, of the emergence of emotions. All emotions, above 0.5 - T-test (U-test) to determine the ratio of total to confirmed emotions	172	- a high agreement overall between the emotions measured by the FaceReader and the researchers' opinions (87%) - difference between male/female expressions - FaceReader is efficient in measuring emotions with over 87% during computer based assessment: successfully recognized Neutral, Surprised, Happy, Scared, and Sad emotions: Limitation in the case of Angry and Disgusted.

	Citation	Scimago ranking	Context	Aim	Methods used (data collection, procedure, statistical analysis methods)	Sample size	Conclusion
3.	Lewinski et al., 2014	Q2	emotions detected while watching amusing video advertisements	Validation: to examine the relationship between the patterns of the FEBE and the perceived amusement of the advertisements, attitude toward the advertisement and attitude toward the brand	Friedman tests are used to determine differences in the global mean average score of the top 10% peak values of facial expressions of happiness	90	FaceReader scores correlated with self-reported effectiveness of the advertising. Only happiness correlated positively with attitude towards the advertisement and the brand in high and medium amusing ads but not low amusing ads.
4.	Lewinski et al., 2014	Q2	calculation with the software on a database	- comparing the performance of FR with human coders (manual facial coding) in terms of AUs and basic emotions - providing an objective accuracy score for FaceReader	- calculation of the agreement index for measuring human inter-coder reliability and comparison of human and computer coding results - Confusion matrix: defining the precision and the accuracy of the recognition - The matching score for FR was the percentage of images that were recognised with the predicted label - FaceReader correctly recognised 89% of expressions on average, whereas human participants only recognised 85%.	52	The software reached a FACS index of agreement of 0.69 on average in 2 datasets: (1) Warsaw Set of Emotional Facial Expression Pictures (WSEFEP) (2) Amsterdam Dynamic Facial Expression Set (ADFES) FaceReader 6.0 has proven to be a reliable indicator of basic emotions, as good at recognising emotions as humans It is less reliable in the case of AU-s: The software achieved an average FACS index of agreement of 0.69 across both datasets.
5.	Chiu et al., 2014	Q1	education: watching multimedia presentations about the physics topic	How learners respond to counterintuitive data and how FMES changes relate to these learners' attempts	What was the dominant expression on their face when they heard the new and important information? Within the students experiencing conceptual change, a significance test between those who had FMES change and	102	Facial microexpression state (FMES) changes can be used to identify moments of conceptual conflict, one of the pathways to conceptual change

	Citation	Scimago ranking	Context	Aim	Methods used (data collection, procedure, statistical analysis methods)	Sample size	Conclusion
				to reconcile the anomalous data.	without: the difference is significant ($z = 1.795, \alpha = .05$, right-tailed test)		
6.	Lewinski, 2015	Q2	Nonverbal communication in social media advertising campaigns	To explore distinct facial expression patterns to forecast the popularity of YouTube videos among social media users.	Facial expressions were categorised as intense (averaged 10% of the highest likelihood and intensity every frame). Comparison of popularity in t2 with facial expressions: mean, standard deviation, and minimum/maximum values for dependent (video views in t2) and independent variables (video views in t1, happiness, sadness, surprise, rage, fear, disgust). - Linear regression, multiple linear regression to check the prediction of t2 views based on t1 views and facial expressions of happiness, sadness, and surprise	16	Successful demonstration of how a neuromarketing approach might be integrated into research on corporate communication. The fewer affiliative emotions (happiness and sadness) that were present, the more popular the video was. However, an increase in surprise increased video views in t2.
7.	Lewinski, 2015	Q2	recognising general emotions based on static photos	Contrasted human recognition scores of 100 typical, neutral front-up facial images with scores of automated facial coding software.	1. Actors posed neutral faces. 2. images coded by a certified FACS coder and by FaceReader 4. Analysis: Independent samples t-test to determine differences in accuracy scores between humans and FaceReader. 5. Levene's test for equality of variances 6. Cohen's effect size 7. Average numbers of human/computer recignition success	100	- the first-ever evidence that computer software (90%)was more accurate in recognising neutral faces than people were(59%). Possible explanations: a) smile-as-a-baseline b) false recognition of emotion - no homogeneity of variances ($p < 0.0005$) - Cohen's effect size value ($d = 1.68$) suggested a high practical significance (Cohen, 1988).
8.	Breaban & Noussair, 2018	Q1	brokers' emotions while working in the stock market	relationship between trader emotions and asset market behaviour, , during work	1-hour-long recordings values compared are the following behaviours: selling, buying, emotions, income		Fear correlates with selling, low prices, and price decreases. The experiment confirms the intuition that emotions and market dynamics are closely related.

	Citation	Scimago ranking	Context	Aim	Methods used (data collection, procedure, statistical analysis methods)	Sample size	Conclusion
9.	Neomániová et al., 2018	Q3	The watched social advertising spot focused on the issue of irresponsible driving (21 mins, 1 spot)	Comparing the results of advertising tests in simulated laboratory conditions with the usage of - EEG device - and FR	Using EEG devices and FR, we tried to reveal emotional state (**valence**) under conditions of different video and audio appeals. - measured emotions in 10-second-long intervals. - calculated average values of the individual emotions of respondents measured during whole video watching: one value for each respondent from which the average and the standard deviation were calculated	16	EEG is a more reliable valence detection tool than FR During video watching, facial expressions detected by FR as neutral, while the results of the EEG indicates prevailing negative emotions in the same time The combination of at least some of them can produce valuable results for marketing management, depending on the character of the research problem.
10.	Hirt et al., 2019	Q2		to check the accuracy/reliability of FaceReader, comparing the results of self-reports and automated facial emotion recognition	-aggregated FaceReader's estimates over each trial; - calculated the mean over 10% of the highest estimates within each trial for boredom and interest - linear mixed-effects regression models - statistical comparison: Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC)	103	FaceReader's aggregated facial emotion recognition estimations did not correlate with self-reports. - Boredom, interest and valence estimations did not predict the self-report measure regardless of face emotion recognition estimate aggregation. Affective computing cannot inevitably model other emotional components based on subjective experience.
11.	Hetland et al., 2019	Q2	Emotions of extreme sport athletes	(1) replicate the findings from Hetland et al. (2018) that happiness arises first after the activity is over, (2) test if the newly developed measure of online interest can provide us with a measure of	Data collection: - During run: emotions measured by FR - After downhill run, questionnaire about state emotions Analysis: - The video is divided into 3 segments (1) before the run, (2) during, and (3) after the run. - expressions examined: interest, AU1 (inner brow raiser) and AU4 (brow lowered) - the mean automatically graded emotion scores were entered in a 8 (emotion: angry, disgusted, happy, interest, neutral, sad, scared, surprised) × 3 (event: before, during, after) ANOVA treating	24	- facially expressed emotions are unrelated to self-reported emotions - Anger was the most prominent of the facially expressed emotions before and during the activity, after it, it decreased (reason: similar to concentration) - doubts about whether interest can be measured according to the current description

	Citation	Scimago ranking	Context	Aim	Methods used (data collection, procedure, statistical analysis methods)	Sample size	Conclusion
				online interest, and finally (3) to explore the relation between measures of online emotions during the activity and self-report emotions given immediately after the activity is over.	both factors as repeated measures. - comparing mean emotions before and during activity and during and after activity. - post hoc procedure: We calculated separate, one-way ANOVAs for each of the emotions and setup ANOVA contrasts such that before and after the downhill event were compared against facially expressed emotions during the activity. - both unadjusted and Bonferroni-adjusted p		
12.	Gudi & Ivan, 2022	n. a. (software white paper)	actors declared interest, confusion, and boredom	to check the accuracy/reliability of custom expressions of FaceReader	F1 measure, discriminate between the AU classifiers, using the formula: 2 * ((precision * recall) / (precision + recall))	20	Participants were asked to declare interest, confusion, and boredom, then compared the videos' automatic analysis to their intent. - Boredom had the lowest detection accuracy of 55%, meaning the software did not "notice" the emotional state in many cases. - The software detected boredom the most accurately of the three emotional states, with 84.6% accuracy and a low false positive rate. Thus, computerised face analysis does not always detect boredom, but it does so with high probability.
13.	Talen & Den Uyl, 2022	Q1	user behavior while using the website	To measure the effect of the complexity of website design and website tasks on the facial expression "anger".	Data collection: - tasks related to website handling and math problem - self-reports related to personality, mood, and an evaluation of the experiment Analysis: metric for the peak expression: the average of the five seconds of the maximal angry score was calculated from the data	75	During the more complex tasks, the facial expression of anger was higher. --> automatically detected facial expression anger could be used to measure usability aspects of a website.

	Citation	Scimago ranking	Context	Aim	Methods used (data collection, procedure, statistical analysis methods)	Sample size	Conclusion
					Effect of website design complexity and task complexity on the expression angry: one-sided Wilcoxon Signed-Rank test Interaction effect of website type: 2 separate between-subject Mann–Whitney U analyses for the easy and difficult search tasks. Effect of math tasks on the facial expression angry: Friedman's ANOVA Subjective complexity and the subjective difficulty of the search task: one-sided Wilcoxon Signed-Ranks test. The completion time of the search task: one-sided Wilcoxon Signed-Ranks test. Subjective difficulty and completion time of the arithmetic tasks: Friedman's ANOVA		
14.	Landmann, 2023	Q1	methodological review	overview of the use and handling of Noldus FaceReader	syslit review & empirical study applying 5 types of analysis methods	29	FaceReader is a reliable tool for measuring emotions. Depending on the data types and the aims of the study, the most reliable analysis methods include Peak Values: Uses top 10% intensity values for strong emotional responses. Overall Mean Values: Averages emotion intensity over time for general trends. Cut-Off Thresholds: Filters values below 0.5 to focus on significant emotions. Maximum Intensity: Extracts highest emotion intensity for extreme reactions assessment.

…

Appendix C: **Analysis of ethical risks and benefits**

Table C1: Potential risks and attempts made to minimise them

	Situation requiring special attention	Risk	robabilit	Measure
1.	A subordinate relationship between the participants (students) and the teachers, which promotes and supports recruitment.	Students may not take part in the observation on a voluntary basis, but under the pressure of the teacher's expectation	High	Teachers will reward students with extra points for participation, but the same number of points can also be earned through an alternative assignment.
2.	Participants (students) evaluated the teachers who presented the videos. Students may report on their experiences of disengagement and shortcomings in the teachers' presentation.	Power asymmetries exist between students and instructors. If instructors could identify students who criticise their presentation or immediacy, this could affect the relationship and potentially influence assessment.	Low	Students' anonymity will be ensured by using codes or nicknames instead of individual names. Educators will have no access to the database or the detailed analysis of the results; access will be limited to a summary of the results only.
3.	Students may report their experiences of disengagement and the shortcomings of teachers' presentations.	Students will not provide sincere opinions because of being afraid of experiencing disadvantages during their studies. This would reduce the validity of the results.	Low	Informed consent ensures that students are informed about the anonymity and the protection of their data and inputs (see the Annex).
4.	Students may report their experiences of disengagement and the shortcomings of teachers' presentations.	Student reports on feedback and negative emotions related to the prepared video presentation may embarrass teachers.	Medi um	Teachers will not have access to the database and the detailed analysis of the results, only to the results summary. Furthermore, the researcher will provide an interpretation of the results in light of previous research.
5.	Students' faces will be video recorded.	Facial video is highly sensitive information and may be misused.	Low	The data is stored on a separate server within the country, in compliance with data protection regulations. The stored data will be accessible only to the 2 researchers involved in the research (see the above-mentioned names), with strict confidentiality. The videos will be stored for 2 months for identification purposes, after which they will be destroyed, and the data will no longer be identifiable. The collected data and their analysis may be published only anonymously and confidentially, in summary form.
6.	The observation will be led by an unknown researcher and will require	Participation in the observation is time-consuming and may	Medi um	Observations took place in a one-on-one setting in an environment as close as possible to their usual learning environment, in one of the university classrooms. The

	approximately 1 hour of participation.	cause anxiety or stress in the case of some students		observation began and ended with introductions and informal conversations. An undisturbed atmosphere, refreshments, and written information ensured participants felt comfortable.

Table C2: Potential benefits and measures planned to maximise them

	Benefit	Probability	Measure
1.	The participation contributes to increased self-knowledge and awareness in the fields of academic emotions, facial expressions, and multimedia tools used in video presentations.	Medium	Upon the participant's request, the researcher conducting the observation will answer questions and provide professional advice on learning.
2.	Participants gain insight into the background and application opportunities of the a facial expression analysis software	High	The researcher will explain and demonstrate some of the data generated by FaceReader, including charts and visualisations.

Appendix D: **Student questionnaires**

Student Questionnaire

Online questionnaire administered via Qualtrics

Bevezetés

Kedves Válaszadó!
Az alábbi kérdőívben az érzelmi állapotával, a megtekintett videóval és az oktatóval kapcsolatos élményeiről kérdezzük.
A kérdőív kitöltése 5-8 percet vesz igénybe.
A válaszadás önkéntes és névtelen.
Az adatfeldolgozáshoz kérjük, adjon meg egy jeligét (egy tetszőleges szót vagy számot, ld. az asztalon levő papírt az ötletekkel). Ezt a jeligét jegyezze meg, mert majd ugyanezt a jeligét kell beírnia a videómegtekintő felületen is a neve helyett - ennek célja az anonimitás biztosítása.

Jelige: _____

Érzelmi állapot – első mérés

Pozitív és Negatív Affektivitás Skála (PANAS). Az alábbiakban felsorolt szavak különböző érzéseket és érzelmeket írnak le. Kérjük, gondosan olvasson el minden jellemzőt, majd jelölje be azt a számot, amely leginkább kifejezi azt, hogy a JELEN PILLANATBAN hogyan érzi magát. Kérjük, ne gondolkozzon sokat a válaszon!
Válaszskála: 1–5. (1 = egyáltalán nem vagy csak alig | 2 = kicsit | 3 = mérsékelten | 4 = eléggé | 5 = nagyon)

Állítás / jellemző	1	2	3	4	5
zaklatott, feldúlt	○	○	○	○	○
ellenséges	○	○	○	○	○
éber	○	○	○	○	○
megszégyenült	○	○	○	○	○
elhivatott	○	○	○	○	○
ideges	○	○	○	○	○
elszánt, határozott	○	○	○	○	○
figyelmes	○	○	○	○	○
aktív, élénk	○	○	○	○	○
félénk	○	○	○	○	○

Érzelmi állapot – második mérés

Pozitív és Negatív Affektivitás Skála (PANAS). Az alábbiakban felsorolt szavak különböző érzéseket és érzelmeket írnak le. Kérjük, gondosan olvasson el minden jellemzőt, majd jelölje be azt a számot, amely leginkább kifejezi azt, hogy a JELEN PILLANATBAN hogyan érzi magát. Kérjük, ne gondolkozzon sokat a válaszon!
Válaszskála: 1–5. (1 = egyáltalán nem vagy csak alig | 2 = kicsit | 3 = mérsékelten | 4 = eléggé | 5 = nagyon)

Állítás / jellemző	1	2	3	4	5
zaklatott, feldúlt	○	○	○	○	○
ellenséges	○	○	○	○	○
éber	○	○	○	○	○
megszégyenült	○	○	○	○	○
elhivatott	○	○	○	○	○
ideges	○	○	○	○	○
elszánt, határozott	○	○	○	○	○
figyelmes	○	○	○	○	○
aktív, élénk	○	○	○	○	○
félénk	○	○	○	○	○

A megtekintett videóval kapcsolatos élmények

Jelölje be azt a számot, amely a leginkább kifejezi az egyetértését az adott állítással, az „egyáltalán nem értek egyet"-től (1) a „teljes mértékben egyetértek"-ig (5).

Állítás / jellemző	1	2	3	4	5
lelkesnek éreztem magam, amikor megnéztem	○	○	○	○	○
izgalmas volt	○	○	○	○	○
a megnézésének köszönhetően feltöltődöm	○	○	○	○	○
témái érdekfeszítőek számomra	○	○	○	○	○
a megnézése élvezetes volt.	○	○	○	○	○
a megnézésétől jól érzem magam	○	○	○	○	○
magával ragad	○	○	○	○	○
az általa tárgyalt témák tetszenek	○	○	○	○	○
mint tapasztalat pozitív érzés számomra	○	○	○	○	○

Tanulási élmény és megértés

Jelölje be azt a számot, amely a leginkább kifejezi az egyetértését az adott állítással, az „egyáltalán nem értek egyet"-től (1) a „teljes mértékben egyetértek"-ig (5).

Állítás / jellemző	1	2	3	4	5
emlékszem arra, amit most tanultam	○	○	○	○	○
úgy érzem, hogy a most tárgyalt témákat megértettem	○	○	○	○	○
értem a gondolatok közötti összefüggéseket	○	○	○	○	○
értem az oktatóvideóban bemutatott tananyagot	○	○	○	○	○
a tudás, amiről tanulunk, hozzáértővé tesz	○	○	○	○	○
a tananyag tartalma hasznos	○	○	○	○	○
tisztában vagyok vele, hogy milyen elvárásoknak kell megfelelnem	○	○	○	○	○
a most tanultak inspiráló kihívást jelentenek számomra	○	○	○	○	○

Az oktató kommunikációja és viselkedése

Jelölje be azt a számot, amely a leginkább kifejezi az egyetértését az adott állítással, az „egyáltalán nem fordul elő"-től (0) a „nagyon gyakran előfordul"-ig (5).

Állítás / jellemző	0	1	2	3	4	5
Személyes példákat használ és saját tapasztalatokat oszt meg az órán.	○	○	○	○	○	○
Kérdez és hozzászólásra bátorítja a diákokat.	○	○	○	○	○	○
A diákok által felhozott témákról elbeszélget, akkor is, ha a téma nem szerepelt az óra tervezett anyagában.	○	○	○	○	○	○
Használ humort az órán.	○	○	○	○	○	○
Név szerint szólít engem.	○	○	○	○	○	○
Óra előtt vagy után beszélget a diákokkal.	○	○	○	○	○	○
A csoportról a "mi" csoportunként beszél, arról beszél, hogy "mit csinálunk".	○	○	○	○	○	○
Megkérdezi a diákokat, mit gondolnak a feladatokról, a határidő vagy a téma kapcsán.	○	○	○	○	○	○
Ha a diákoknak kérdése van, vagy meg akarnak beszélni valamit, akkor kész konzultálni az órán kívül is.	○	○	○	○	○	○
Olyan kérdéseket tesz fel, amelyek nézőpontokra vagy véleményekre kérdeznek rá.	○	○	○	○	○	○
Megdicséri a diákok munkáját, cselekedeteit vagy megjegyzéseit.	○	○	○	○	○	○
Magabiztosan kezeli a technológiai eszközöket.	○	○	○	○	○	○

Original English Items Corresponding to the Hungarian Questionnaire

Note. The questionnaire was administered in Hungarian via Qualtrics. For established measures, the original English wording of the corresponding items is reproduced below rather than back-translating the Hungarian items. Where wording was adapted to the instructional-video context or where a study-specific item was added, this is indicated in the accompanying notes.

Student Questionnaire

Online questionnaire administered via Qualtrics

Introduction

Dear Participant,

The following questionnaire asks about your current emotional state and your experiences related to the video you have watched and the instructor.

The questionnaire takes approximately 5-8 minutes to complete.

Participation is voluntary and anonymous.

For data processing purposes, please provide a code name (any word or number; see the sheet on the desk for suggestions). Please remember this code name, as you will be asked to enter the same code instead of your name on the video-viewing platform. This is intended to ensure anonymity.

Code name: _____

Emotional state - first measurement

Positive and Negative Affect Schedule (PANAS), short form. The words listed below describe different feelings and emotions. Please read each item carefully and select the number that best describes how you feel RIGHT NOW. Please do not spend too much time thinking about your response.

Response scale: 1-5 (1 = very slightly or not at all | 2 = a little | 3 = moderately | 4 = quite a bit | 5 = very much)

Item / descriptor	1	2	3	4	5
upset	○	○	○	○	○
hostile	○	○	○	○	○
alert	○	○	○	○	○
ashamed	○	○	○	○	○
inspired	○	○	○	○	○
nervous	○	○	○	○	○
determined	○	○	○	○	○
attentive	○	○	○	○	○
active	○	○	○	○	○
afraid	○	○	○	○	○

Emotional state - second measurement

Positive and Negative Affect Schedule (PANAS), short form. The words listed below describe different feelings and emotions. Please read each item carefully and select the number that best describes how you feel RIGHT NOW. Please do not spend too much time thinking about your response.

Response scale: 1-5 (1 = very slightly or not at all | 2 = a little | 3 = moderately | 4 = quite a bit | 5 = very much)

Item / descriptor	1	2	3	4	5
upset	○	○	○	○	○
hostile	○	○	○	○	○
alert	○	○	○	○	○
ashamed	○	○	○	○	○
inspired	○	○	○	○	○
nervous	○	○	○	○	○

Item / descriptor	1	2	3	4	5
determined	○	○	○	○	○
attentive	○	○	○	○	○
active	○	○	○	○	○
afraid	○	○	○	○	○

Experiences related to the video

Please select the number that best represents your agreement with each statement, from strongly disagree (1) to strongly agree (5).

Student Interest Scale - Emotional Interest

The original Mazer (2012) items use the stem: "I am interested in this class because ..." In the administered Hungarian questionnaire, the wording was adapted to refer to the instructional video.

Original English item	1	2	3	4	5
I feel enthused about being in class.	○	○	○	○	○
The class makes me feel excited.	○	○	○	○	○
The class causes me to feel energized.	○	○	○	○	○
The topics covered in the course fascinate me.	○	○	○	○	○
Being in the class is enjoyable.	○	○	○	○	○
The class experience makes me feel good.	○	○	○	○	○
The material fascinates me.	○	○	○	○	○
I like the things we cover in class.	○	○	○	○	○
The class experience feels very positive.	○	○	○	○	○

Learning experience and understanding

Please select the number that best represents your agreement with each statement, from strongly disagree (1) to strongly agree (5).

Student Interest Scale - Cognitive Interest

The original Mazer (2012) items use the stem: "I am interested in this class because ..." In the administered Hungarian questionnaire, the wording was adapted to refer to the learning material and the instructional video.

Original English item	1	2	3	4	5
I can remember the course material.	○	○	○	○	○
I feel like I am learning topics covered in the course.	○	○	○	○	○
I can understand the flow of ideas.	○	○	○	○	○
I understand the course material.	○	○	○	○	○
The information covered in the course is making me more knowledgeable.	○	○	○	○	○
The information in the course is useful.	○	○	○	○	○
I realize what is expected of me.	○	○	○	○	○

Additional study-specific item: What I have just learned presents an inspiring challenge to me.

Instructor communication and behaviour

Please indicate how frequently each behaviour occurred. In the present questionnaire, responses were recorded on a 0-5 frequency scale.

Verbal Teacher Immediacy - adapted from Gorham (1988)

Original English item	0	1	2	3	4	5
Uses personal examples or talks about experiences she/he has h outside of class.	○	○	○	○	○	○
Asks questions or encourages students to talk.	○	○	○	○	○	○
Gets into discussions based on something a student brings up ev when this does not seem to be part of his/her lecture plan.	○	○	○	○	○	○
Uses humor in class.	○	○	○	○	○	○
Addresses me by name.	○	○	○	○	○	○
Gets into conversations with individual students before or af class.	○	○	○	○	○	○
Refers to class as "our" class or what "we" are doing.	○	○	○	○	○	○

Original English item	0	1	2	3	4	5
Asks how students feel about an assignment, due date or discuss topic.	○	○	○	○	○	○
Invites students to telephone or meet with him/her outside of cl if they have questions or want to discuss something.	○	○	○	○	○	○
Asks questions that solicit viewpoints or opinions.	○	○	○	○	○	○
Praises students' work, actions or comments.	○	○	○	○	○	○

Additional study-specific item: Uses technological equipment confidently.

Measurement Instruments and Scoring Procedures

Where established measures were adapted to the study's specific context, only items applicable and meaningful in the video-based instructional setting were retained. The tables below therefore report both the structure of the original instruments and the items actually used in the present study; reliability estimates were calculated for the latter.

Measure / variable	Scale and scoring	Naming in the thesis
Student Interest Scale (SIS) – Mazer (2012)	16 items rated on a 5-point Likert scale from *strongly disagree* to *strongly agree*. The scale comprises two subscales: Emotional Interest (EI; 9 items) and Cognitive Interest (CI; 7 items). Mazer reports summed scores for the two subscales; therefore, possible scores range from 9–45 for EI and 7–35 for CI. Higher scores indicate higher levels of the respective form of interest.	Emotional Interest (EI) and Cognitive Interest (CI)
Positive and Negative Affect Schedule (PANAS) – Watson, Clark and Tellegen (1988); Hungarian validation: Gyollai et al. 2011	20 items rated on a 5-point Likert-type scale. Ten items form the Positive Affect (PA) subscale and ten form the Negative Affect (NA) subscale. Item scores are summed separately, producing PA and NA scores ranging from 10–50. Higher values indicate higher levels of positive or negative affect, respectively. PA and NA are treated as relatively independent dimensions, rather than opposite ends of a single continuum.	Positive Affect (PA) and Negative Affect (NA). If measured before and after the video: PA-pre, PA-post, NA-pre, NA-post. Change scores may be labelled ΔPA and ΔNA.
Verbal Immediacy Scale – Gorham, (1988)	In the original 14-item measure, respondents first indicated whether each behaviour occurred. If present, frequency was rated from 1 = rarely to 4 = very often; absence of the behaviour was scored 0. After reverse-scoring the non-immediate items where necessary, item scores were summed to produce a total immediacy score. Higher scores indicate greater perceived teacher immediacy.	Nonverbal Immediacy or Total Immediacy Score

The validated Hungarian version of the original 20-item PANAS was used (Gyollai et al., 2011). Confirmatory factor analysis supported the theoretical two-factor structure of Positive and Negative Affect, and internal consistency was above .80 for both original subscales.

The Student Interest Scale was developed and validated by Mazer (2012). Exploratory and confirmatory factor analyses supported a two-factor structure comprising Emotional Interest and Cognitive Interest, with high internal consistency reported for both subscales.

Reliability table on internal consistency

Measure/subscale	Items in original	Items used	Cronbach's α
Emotional Interest Scale	9	9	.93
Cognitive Interest Scale	7	7	.84
Verbal Immediacy Scale	20	11	.92
Positive Affectivity Scale pre-test	10	5	.79
Negative Affectivity Scale pre-test	10	5	.46
Positive Affectivity Scale post-test	10	5	.91
Negative Affectivity Scale post-test	10	5	.69

References

Gorham, J. (1988). *The relationship between verbal teacher immediacy behaviors and student learning*. Communication Education, 37(1), 40–53.

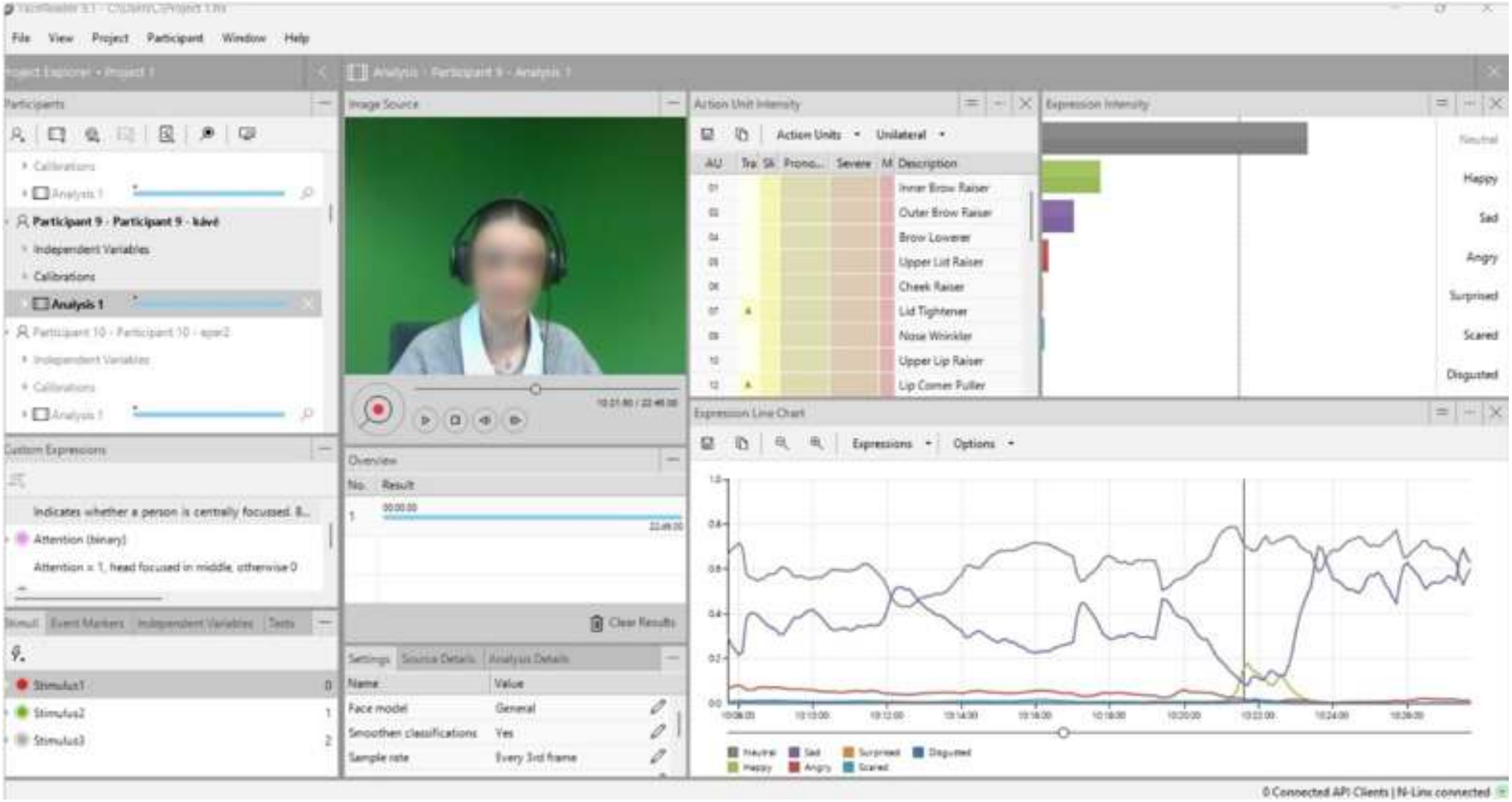
Gyollai, Á., Simor, P., Köteles, F. and Demetrovics, Z. (2011) 'Psychometric properties of the Hungarian version of the original and the short form of the Positive and Negative Affect Schedule (PANAS)', *Neuropsychopharmacologia Hungarica*, 13(2), pp. 73–79.

Mazer, J.P. (2012) 'Development and validation of the Student Interest and Engagement Scales', *Communication Methods and Measures*, 6(2), pp. 99–125. doi: 10.1080/19312458.2012.679244.

Watson, D. and Tellegen, A. (1985) 'Toward a consensual structure of mood', *Psychological Bulletin*, 98(2), pp. 219–235.

Watson, D., Clark, L.A. and Tellegen, A. (1988) 'Development and validation of brief measures of positive and negative affect: The PANAS scales', *Journal of Personality and Social Psychology*, 54(6), pp. 1063–1070.

Appendix E: Screenshot of the FaceReader 9.1 interface (Noldus Information Technology)



Source: Own capture.

Appendix F: Effects of multimedia learning features on facial expressions related to boredom

Table 1: Effect of applied multimedia principles on facial expression “neutral” as identified by FaceReader, using the Wilcoxon signed rank test and Student t-test (JASP (Version 0.19.2)).

Paired Samples T-Test

												95% CI for Effect Size	
Measure 1		Measure 2	Test	Statistic	z	df	p	Rank	Holm-Bonferroni corrected p	Adjusted significance	Effect Size	SE Effect Size	Lower	Upper
Neutral | No visuals	-	Neutral | Visual element	Wilcoxon	1065	4.127		0.000038	1	0.00041	Significant	0.671	0.161	0.457	0.811
Neutral | Radundant	-	Neutral | Not redundant	Wilcoxon	990	3.403		0.000679	2	0.00679	Significant	0.553	0.161	0.295	0.736
Neutral | No signalling	-	Neutral | Signalling	Wilcoxon	947	2.988		0.003000	3	0.02700	Significant	0.485	0.161	0.209	0.69
Neutral | No humour	-	Neutral | Humour	Wilcoxon	910	2.631		0.009000	4	0.07200	Non significant	0.427	0.161	0.137	0.65
Neutral | No Smile	-	Neutral | Smile	Student	2.54		49	0.014000	5	0.09800	Non significant	0.359	0.071	0.071	0.644
Neutral | No face	-	Neutral | Face	Wilcoxon	676	2.302		0.022000	6	0.13200	Non significant	0.374	0.161	0.075	0.612
Neutral | Not Coherent	-	Neutral | Coherent	Student	2.242		49	0.029000	7	0.14500	Non significant	0.317	0.038	0.031	0.6
Neutral | No personalization	-	Neutral | Personal	Student	2.018		49	0.049000	8	0.19600	Non significant	0.285	0.033	0.001	0.567
Neutral | No example	-	Neutral | Example	Wilcoxon	778	1.356		0.177000	9	0.53100	Non significant	0.22	0.161	-0.094	0.495
Neutral | Slide rest	-	Neutral | Slide first minute	Wilcoxon	621	-0.159		0.877000	10	1.00000	Non significant	-0.026	0.161	-0.331	0.285
Neutral | No story	-	Neutral | Story	Wilcoxon	653	0.15		0.885000	11	0.88500	Non significant	0.024	0.161	-0.286	0.33

e Student t-test, the effect size is given by Cohen's *d*. For the Wilcoxon test, the effect size is given by the matched rank biserial correlation.

ed cells with grey indicate large (r_{rb} > 0.5) effect size, as defined by (Cureton, 1956) for the Wilcoxon Signed-Rank Test. The lightest grey indicates non-significant but noteworthy results (p ≥ 0.05).

Table 2: Effect of certain multimedia principles on “valence” and “arousal” as measured by FaceReader, using Student t-test (JASP (Version 0.19.2).

Paired Samples Student T-Test (One-tailed)

Measure 1		Measure 2	t	df	p	Rank	Holm-Bonferro corrected p	Adjusted Significance	Cohen's d	SE Cohen's d
Valence | No personalisation	-	Valence | Personal	-6.153	49	0.00000007	1	2.7168E-07	Significant	-0.870	0.047
Arousal | No personalisation	-	Arousal | Personal	-5.961	49	0.00000013	2	4.032E-07	Significant	-0.843	0.059
Valence | Slide rest	-	Valence | Slide first minute	-2.084	49	0.02100000	3	0.042	Significant	-0.295	0.031
Arousal | No story	-	Arousal | Story	-0.998	49	0.16200000	4	0.162	Not Significant	-0.141	0.067

Note. For all tests, the alternative hypothesis specifies that Measure 1 is less than Measure 2. For example, Valence | No personalization is less than Valence | Personal.
Note. Shaded cells with grey indicate small (d > 0.2), medium (d > 0.5), and large (d > 0.8) effect sizes, as defined by Cohen (1988).

Table 3: Effect of certain multimedia tools on “arousal” as measured by FaceReader, using the Wilcoxon signed rank test (JASP (Version 0.19.2)).

Wilcoxon signed rank test (One-tailed)

Measure 1	Measure 2	W	z	P		Holm-Bonferroni corrected p	Adjusted Significance	Effect size	t size	95% CI for Effect size Lower	Upper
Arousal | No humour	Arousal | Humour	204	-4.185	0.00001	1	0.00013	Significant	-0.680	0.161	-∞	-0.509
Arousal | No face	Arousal | Face	226	-3.972	0.00004	2	0.00029	Significant	-0.645	0.161	-∞	-0.462
Arousal | Slide rest	Arousal | Slide first minute	316	-3.104	0.00097	3	0.00680	Significant	-0.504	0.161	-∞	-0.28
Arousal | No visuals	Arousal | Visual element	379	-2.495	0.00600	4	0.03600	Significant	-0.405	0.161	-∞	-0.161
Arousal | Redundant	Arousal | Not redundant	439	-1.916	0.02800	5	0.14000	Non-Significant	-0.311	0.161	-∞	-0.055
Arousal | No Smile	Arousal | Smile	442	-1.887	0.03000	6	0.12000	Non-Significant	-0.307	0.161	-∞	-0.05
Arousal | No example	Arousal | Example	745	1.038	0.85100	7	1.00000	Non-Significant	0.169	0.161	-∞	0.412
Arousal | No signalling	Arousal | Signalling	854	2.09	0.98200	8	1.00000	Non-Significant	0.340	0.161	-∞	0.552
Arousal | Not Coherent	Arousal | Coherent	263	-3.615	1.00000	9	1.00000	Non-Significant	-0.587	0.161	-∞	-0.385

Note. For all tests, the alternative hypothesis specifies that Measure 1 is less than Measure 2. For example, Arousal | No humour is less than Arousal | Humour.
Note. Effect size is given by the matched rank-biserial correlation.
Note. Shaded cells with grey indicate small (r_{rb} > 0.1), medium (r_{rb} > 0.3) and large (r_{rb} > 0.5) effect sizes, as defined by (Cureton, 1956). The lightest grey indicates non-significant but noteworthy results (p ≥ 0.05).

Appendix G: Effects of multimedia learning features on facial expressions related to basic emotions

Note: Cells highlighted with 3 shades of grey indicate small ($r_{rb} > 0.1$), medium ($r_{rb} > 0.3$) and large ($r_{rb} > 0.5$) effect sizes, as defined b (Cureton, 1956) for the Wilcoxon Signed-Rank Test.

Table 1: Effect of certain multimedia principles on facial expression "happy" as identified by FaceReader, using Wilcoxon signed rank test (JASP (Version 0.19.2)).

Wilcoxon signed-rank test

							95% CI for Rank-Biserial Correlation	
Measure 1	Measure 2	W	z	p	Rank-Biserial Correlation	SE Rank-Biserial Correlation	Lower	Upper
Happy | No face	Happy | Face	84	-5,343	< .001	-0,868	0,161	-0,928	-0,765
Happy | No humour	Happy | Humour	106	-5,131	< .001	-0,834	0,161	-0,908	-0,707
Happy | No personalization	Happy | Personal	120	-4,996	< .001	-0,812	0,161	-0,896	-0,672
Happy | Redundant	Happy | Not redundant	120	-4,996	< .001	-0,812	0,161	-0,896	-0,672
Happy | No Smile	Happy | Smile	142	-4,783	< .001	-0,777	0,161	-0,876	-0,617
Happy | Not Coherent	Happy | Coherent	176	-4,455	< .001	-0,724	0,161	-0,844	-0,535
Happy | No visuals	Happy | Visual element	237	-3,866	< .001	-0,628	0,161	-0,785	-0,397
Happy | No example	Happy | Example	283	-3,422	< .001	-0,556	0,161	-0,738	-0,299
Happy | No story	Happy | Story	658	0,198	0,847	0,032	0,161	-0,279	0,337
Happy | Slide rest	Happy | Slide first minute	658	0,198	0,847	0,032	0,161	-0,279	0,337
Happy | No signalling	Happy | Signalling	1098	4,445	< .001	0,722	0,161	0,533	0,843

Note. Wilcoxon signed-rank test

Note. Coloured cells indicate large effect sizes ($r_{rb} > 0.5$), as defined by Cureton (1956), while non-significant results ($p \geq 0.05$) have been excluded from categorisation

Table 2: Effect of certain multimedia principles on "sad" as measured by FaceReader, using Wilcoxon signed rank test (JASP (Version 0.19.2)).

Wilcoxon signed rank test (Two-tailed)

										95% CI for Effect Size	
Measure 1	Measure 2	W	z	p	Rank	Holm-Bonferroni corrected p	Adjusted significance	Effect Size	SE Effect Size	Lower	Upper
Sad | No Smile	Sad | Smile	1021	4.063	0.00004938	1	0.000543	Significant	0.667	0.163	0.449	0.81
Sad | No signalling	Sad | Signalling	226	-3.972	0.00007263	2	0.000726	Significant	-0.645	0.161	-0.795	-0.421
Sad | No personalization	Sad | Personal	1048	3.963	0.00007563	3	0.000681	Significant	0.644	0.161	0.419	0.794
Sad | No humour	Sad | Humour	978	3.287	0.001	4	0.008000	Significant	0.534	0.161	0.27	0.723
Sad | Slide rest	Sad | Slide first minute	923	2.756	0.006	5	0.042000	Significant	0.448	0.161	0.162	0.664
Sad | No face	Sad | Face	900	2.534	0.011	6	0.066000	Non-Significant	0.412	0.161	0.119	0.639
Sad | No story	Sad | Story	843	1.984	0.048	7	0.240000	Non-Significant	0.322	0.161	0.016	0.574
Sad | Redundant	Sad | Not redundant	464	-1.675	0.095	8	0.380000	Non-Significant	-0.272	0.161	-0.535	0.039
Sad | No example	Sad | Example	776	1.337	0.183	9	0.549000	Non-Significant	0.217	0.161	-0.097	0.492
Sad | Not Coherent	Sad | Coherent	773	1.308	0.193	10	0.386000	Non-Significant	0.213	0.161	-0.102	0.489
Sad | No visuals	Sad | Visual element	525	-1.086	0.28	11	0.280000	Non-Significant	-0.176	0.161	-0.46	0.139

Note. For the Wilcoxon test, the effect size is given by the matched rank biserial correlation.

Note. Shaded cells with grey indicate small ($r_{rb} > 0.1$), medium ($r_{rb} > 0.3$) and large ($r_{rb} > 0.5$) effect sizes, as defined by (Cureton, 1956) The lightest grey indicates non-significant but noteworthy results ($p \neq 0.05$)

Table 3: Effect of certain multimedia principles on “angry” as measured by FaceReader, using Wilcoxon signed rank test (JASP (Version 0.19.2)).

Wilcoxon signed rank test (Two-tailed)

Measure 1		Measure 2	W	z	p	Rank	Holm-Bonferroni corrected p	Adjusted significance	Effect Size	SE Effect Size	95% CI for Effect Size	
											Lower	Upper
Angry | No humour	-	Angry | Humour	1063	4.301	0.00001741	1	0.00019151	Significant	0.689	0.161	0.486	0.829
Angry | No signaling	-	Angry | Signaling	225	-4.001	0.0000643	2	0.000643	Significant	-0.65	0.161	-0.798	-0.425
Angry | No example	-	Angry | Example	665	3.355	0.000809	3	0.007281	Significant	0.545	0.161	0.285	0.731
Angry | No story	-	Angry | Story	964	3.152	0.002	4	0.016	Significant	0.512	0.161	0.242	0.706
Angry | No visuals	-	Angry | Visual element	339	-2.882	0.004	5	0.028	Significant	-0.468	0.161	-0.678	-0.187
Angry | No personalization	-	Angry | Personal	922	2.746	0.006	6	0.036	Significant	0.446	0.161	0.16	0.663
Angry | No Smile	-	Angry | Smile	915	2.679	0.007	7	0.039	Significant	0.435	0.161	0.147	0.666
Angry | No face	-	Angry | Face	850	2.051	0.041	8	0.164	Non significant	0.333	0.161	0.028	0.582
Angry | Side rest	-	Angry | Side first minute	440	-1.907	0.057	9	0.171	Non significant	-0.31	0.161	-0.564	-0.052
Angry | Redundant	-	Angry | Not redundant	506	-1.25	0.213	10	0.426	Non significant	-0.203	0.161	-0.461	0.112
Angry | Not Coherent	-	Angry | Coherent	554	-0.806	0.423	11	0.423	Non significant	-0.131	0.161	-0.422	0.165

Note. For the Wilcoxon test, the effect size is given by the matched rank biserial correlation.

Note. Shaded cells with grey indicate small (r_{rb} > 0.1), medium (r_{rb} > 0.3) and large (r_{rb} > 0.5) effect sizes, as defined by (Cureton, 1956). The lightest grey indicates non-significant but noteworthy results (p < 0.05).

Table 4: Effect of certain multimedia principles on “surprised” as measured by FaceReader, using Wilcoxon signed rank test (JASP (Version 0.19.2)).

Wilcoxon signed rank test (Two-tailed)

Measure 1		Measure 2	W	z	p	Rank	Holm-Bonferroni	Adjusted significance	Effect Size	SE Effect Size	95% CI for Effect Size	
											Lower	Upper
Surprised | No Smile	-	Surprised | Smile	856	2.109	0.035	1	0.385	Non Significant	0.343	0.161	0.039	0.589
Surprised | Redundant	-	Surprised | Not redundant	484	-1.482	0.140	2	1.000	Non Significant	-0.241	0.161	-0.511	0.073
Surprised | No signalling	-	Surprised | Signalling	769	1.269	0.206	3	1.000	Non Significant	0.206	0.161	-0.109	0.484
Surprised | Not Coherent	-	Surprised | Coherent	509	-1.240	0.217	4	1.000	Non Significant	-0.202	0.161	-0.480	0.114
Surprised | Side rest	-	Surprised | Side first minute	548	-0.864	0.390	5	1.000	Non Significant	-0.140	0.161	-0.430	0.175
Surprised | No example	-	Surprised | Example	689	0.761	0.450	6	1.000	Non Significant	0.125	0.163	-0.194	0.420
Surprised | No personalization	-	Surprised | Personal	593	-0.430	0.671	7	1.000	Non Significant	-0.070	0.161	-0.370	0.244
Surprised | No face	-	Surprised | Face	665	0.265	0.794	8	1.000	Non Significant	0.043	0.161	-0.269	0.347
Surprised | No visuals	-	Surprised | Visual element	629	-0.082	0.938	9	1.000	Non Significant	-0.013	0.161	-0.320	0.296
Surprised | No story	-	Surprised | Story	644	0.063	0.954	10	1.000	Non Significant	0.010	0.161	-0.299	0.317
Surprised | No humour	-	Surprised | Humour	644	0.063	0.954	11	0.954	Non Significant	0.010	0.161	-0.299	0.317

Note. For the Wilcoxon test, the effect size is given by the matched rank biserial correlation.

Note. Shaded cells with grey indicate small (r_{rb} > 0.1), medium (r_{rb} > 0.3) and large (r_{rb} > 0.5) effect sizes, as defined by (Cureton, 1956). The lightest grey indicates non-significant but noteworthy results (p > 0.05).

Table 5: Effect of certain multimedia principles on “scared” as measured by FaceReader, using Wilcoxon signed rank test (JASP (Version 0.19.2)).

Wilcoxon signed rank test (Two-tailed)

Measure 1		Measure 2	W	z	p	Rank	Holm-Bonferroni corrected p	Adjusted significance	Effect Size	SE Effect Size	95% CI for Effect Size	
											Lower	Upper
Scared | No signalling	-	Scared | Signalling	438.5	-1.921	0.055	1	0.6050	Non significant	-0.312	0.161	-0.566	-0.004
Scared | No face	-	Scared | Face	662.0	1.588	0.113	2	1.0000	Non significant	0.259	0.161	-0.054	0.524
Scared | No humour	-	Scared | Humour	782.0	1.395	0.165	3	1.0000	Non significant	0.227	0.161	-0.088	0.5
Scared | No Smile	-	Scared | Smile	768.5	1.265	0.205	4	1.0000	Non significant	0.205	0.161	-0.11	0.483
Scared | No example	-	Scared | Example	762.0	1.202	0.231	5	1.0000	Non significant	0.195	0.161	-0.12	0.475
Scared | No visuals	-	Scared | Visual element	533.0	-1.009	0.315	6	1.0000	Non significant	-0.164	0.161	-0.449	0.152
Scared | No story	-	Scared | Story	703.0	0.900	0.371	7	1.0000	Non significant	0.148	0.163	-0.171	0.439
Scared | Not Coherent	-	Scared | Coherent	574.0	-0.613	0.543	8	1.0000	Non significant	-0.100	0.161	-0.396	0.215
Scared | Slide rest	-	Scared | Slide first minute	531.0	-0.585	0.562	9	1.0000	Non significant	-0.097	0.164	-0.399	0.224
Scared | No personalization	-	Scared | Personal	664.5	0.517	0.608	10	1.0000	Non significant	0.083	0.163	-0.232	0.386
Scared | Redundant	-	Scared | Not redundant	257.0	-3.673	0.00024	11	0.0002442	Significant	-0.597	0.161	-0.764	-0.354

Note. For the Wilcoxon test, the effect size is given by the matched rank biserial correlation.

Note. Shaded cells with grey indicate small (r_{rb} > 0.1), medium (r_{rb} > 0.3) and large (r_{rb} > 0.5) effect sizes, as defined by (Cureton, 1956). The lightest grey indicates non-significant but noteworthy results (p ≥ 0.05).

Table 6: Effect of certain multimedia principles on “disgusted” as measured by FaceReader, using Wilcoxon signed rank test (JASP (Version 0.19.2)).

Wilcoxon signed rank test (Two-tailed)

Measure 1		Measure 2	W	z	p	Rank	Holm-Bonferroni corrected p	Adjusted significance	Effect Size	SE Effect Size	95% CI for Effect Size	
											Lower	Upper
Disgusted | No visuals	-	Disgusted | Visual element	278.0	-3.470	0.000529	1	0.00582	Significant	-0.564	0.161	-0.743	-0.310
Disgusted | Redundant	-	Disgusted | Not redundant	301.0	-3.248	0.001000	2	0.01000	Significant	-0.528	0.161	-0.719	-0.262
Disgusted | No face	-	Disgusted | Face	330.0	-2.968	0.003	3	0.02700	Significant	-0.482	0.161	-0.688	-0.205
Disgusted | Not Coherent	-	Disgusted | Coherent	362.0	-2.466	0.014	4	0.11200	Non significant	-0.401	0.161	-0.631	-0.106
Disgusted | No Smile	-	Disgusted | Smile	446.0	-1.849	0.065	5	0.45500	Non significant	-0.300	0.161	-0.557	0.009
Disgusted | No personalization	-	Disgusted | Personal	451.0	-1.800	0.073	6	0.43800	Non significant	-0.293	0.161	-0.551	0.017
Disgusted | No signalling	-	Disgusted | Signalling	813.0	1.694	0.091	7	0.45500	Non significant	0.275	0.161	-0.036	0.538
Disgusted | No example	-	Disgusted | Example	494.5	-1.380	0.169	8	0.67600	Non significant	-0.224	0.161	-0.486	0.090
Disgusted | No story	-	Disgusted | Story	515.0	-1.183	0.239	9	0.71700	Non significant	-0.192	0.161	-0.472	0.123
Disgusted | No humour	-	Disgusted | Humour	516.0	-1.173	0.243	10	0.48600	Non significant	-0.191	0.161	-0.471	0.125
Disgusted | Slide rest	-	Disgusted | Slide first minute	540.0	-0.941	0.349	11	0.34900	Non significant	-0.153	0.161	-0.440	0.163

Note. For the Wilcoxon test, the effect size is given by the matched rank biserial correlation.

Note. Shaded cells with grey indicate small (r_{rb} > 0.1), medium (r_{rb} > 0.3) and large (r_{rb} > 0.5) effect sizes, as defined by (Cureton, 1956). The lightest grey indicates non-significant but noteworthy results (Holm-Bonferroni corrected p ≥ 0.05).

Appendix H: Interview questions

The retrospective interviews were semi-structured and were organised around two main questions. The wording and sequence of the questions varied somewhat across participants, and follow-up questions were used where appropriate to clarify or elaborate on participants' responses.

The two main questions were:

1. Which parts of the video did you find particularly easy or interesting to follow?
2. Which parts of the video did you find more difficult to follow, or where did you feel that your attention declined?

Follow-up questions explored which features of the video contributed to these experiences, with more specific probes used where relevant. These included features that participants perceived as supporting or impairing their attention, interest, and understanding.

Data availability. Anonymised interview transcripts are not included in the dissertation but may be made available for research verification upon reasonable request, subject to the ethical and data-protection conditions applicable to the study.