Video tutorials in traditional classrooms: The effects on cognitive load and time on task

Dissertation by Publications

for the academic degree
Doctor of Philosophy (Dr.Phil.)

at the
University of Hamburg
Faculty of Education
Vocational and Business Education

submit by
Enqi Fan

Hamburg
2025

Information on the Dissertation Defense

Date of Disputation (Datum der Disputation):
14 October 2025

Written Reviewers (Schriftliche Gutachter:innen):
Prof. Dr. Jens Siemon
Prof. Dr. Kira Weber
Prof. Dr. Jürgen Seifried (University of Mannheim)

Oral Examiner (Mündliche Gutachterin):
Prof. Dr. Freydis Vogel

TABLE OF CONTENTS

Abstract

This cumulative dissertation comprises three papers, each examining the effectiveness of video tutorials (VTs) in the traditional classroom on different types of cognitive load and time-on-task. Each article explores specific research questions related to the use of video tutorials in the classroom and is already published in a peer-review journal or currently under review.

Article Title	Authors	Publication Information
Video Tutorials in the Traditional Classroom: The Effects on Different Types of Cognitive Load	Enqi Fan Matt Bower Jens Siemon	Published: June 2024 *Technology, Knowledge and Learning*
Comparing cognitive load during video versus traditional classroom instruction based on heart-rate variability measures	Enqi Fan Matt Bower Jens Siemon	Submitted: July 2024 (2nd revision Under Review) *Computers & Education*
From Heartbeats to Actions: Multimodal Learning Analytics of Cognitive and Behavior Engagement in Real Classrooms	Enqi Fan Matt Bower Jens Siemon	Submitted: December 2024 (Under Review) *Learning and Instruction*

Each article explores a specific research question relevant to the study. The contributions of each author are listed below:

- Enqi Fan: conceptualization and design of the study, literature review, development of research instruments, collection of data, data analysis, initial drafting of the manuscript, and final editing of the manuscript.
- Matt Bower: provide suggestions for improving the manuscript.
- Jens Siemon: support for the conception and design of the study, support for data collection, support for data analysis, especially database processing, review of manuscripts and providing advice.

Article 1 examines the effects of video tutorials on intrinsic, extraneous, and germane cognitive load. The study was conducted with 45 students in two vocational schools and one high school. The results found that video tutorials significantly reduced intrinsic cognitive load and increased germane cognitive load compared to traditional instruction. These results suggest that video tutorials facilitate deeper learning by optimizing cognitive resource allocation.

Article 2 compares cognitive load during video versus traditional classroom instruction using heart rate variability (HRV) measures, specifically the RMSSD indicator. The study, which surveyed 45 students from two vocational schools and one high school, showed that students experienced higher cognitive load during the developmental phase when using video tutorials. This increased cognitive load is associated with higher engagement and deeper processing of learning material.

Article 3 explores the dynamic relationship between student behavior and cognition through time-on-task and heart rate variability (HRV) data. An analysis of 45 students from two vocational schools and one high school found a significant negative correlation between time-on-task and HRV. In classrooms using videos, students' cognitive changes were more stable. In contrast, students in traditional classrooms showed greater cognitive fluctuations.

The combined results of my three studies demonstrate the potential benefits of using video tutorials in traditional classrooms. The results showed that video tutorials can optimize cognitive load, increase student engagement, and potentially improve learning outcomes. However, there are some limitations, such as a small sample size, individual differences among participants, and insufficient control of variables. In future research these limitations should be solved, and video tutorials should be studied in more depth.

Zusammenfassung

Diese kumulative Dissertation umfasst drei Beiträge, die jeweils die Wirksamkeit von Video-Tutorials (VTs) im traditionellen Unterricht in Bezug auf verschiedene Arten kognitiver Belastung sowie die aufgewendete Lernzeit untersuchen. Jeder Beitrag widmet sich spezifischen Forschungsfragen zum Einsatz von Video-Tutorials im Klassenzimmer und ist entweder bereits in einer peer-reviewten Fachzeitschrift veröffentlicht oder befindet sich derzeit im Begutachtungsverfahren.

Titel des Artikels	**Autor:innen**	**Publikationsinformationen**
Video Tutorials in the Traditional Classroom: The Effects on Different Types of Cognitive Load	Enqi Fan Matt Bower Jens Siemon	Veröffentlicht: Juni 2024 *Technology, Knowledge and Learning*
Comparing cognitive load during video versus traditional classroom instruction based on heart-rate variability measures	Enqi Fan Matt Bower Jens Siemon	Eingereicht: Juli 2024 (2. Revision, unter Begutachtung) *Computers & Education*
From Heartbeats to Actions: Multimodal Learning Analytics of Cognitive and Behavior Engagement in Real Classrooms	Enqi Fan Matt Bower Jens Siemon	Eingereicht: Dezember 2024 (unter Begutachtung) *Learning and Instruction*

Jeder Beitrag befasst sich mit einer spezifischen, forschungsrelevanten Fragestellung im Rahmen der Studie. Die Beiträge der einzelnen Autor:innen sind wie folgt:
Enqi Fan: Konzeption und Design der Studie, Literaturrecherche, Entwicklung der Erhebungsinstrumente, Datenerhebung, Datenanalyse, Erstentwurf des Manuskripts sowie abschließende Redaktion des Manuskripts.
Matt Bower: Fachliche Rückmeldungen und Verbesserungsvorschläge zum Manuskript.
Jens Siemon: Unterstützung bei der Konzeption und dem Design der Studie, bei der Datenerhebung und -analyse (insbesondere bei der Datenbankverarbeitung), sowie Durchsicht der Manuskripte und methodische Beratung.

Artikel 1 untersucht die Auswirkungen von Video-Tutorials auf die intrinsische, extrinsische und lernbezogene (germane) kognitive Belastung. Die Studie wurde mit 45 Schüler:innen an zwei berufsbildenden Schulen und einem Gymnasium durchgeführt. Die Ergebnisse zeigen, dass Video-Tutorials die intrinsische kognitive Belastung signifikant verringern und die lernbezogene Belastung im Vergleich zum traditionellen Unterricht erhöhen. Dies deutet darauf hin, dass Video-Tutorials durch eine optimierte Allokation kognitiver Ressourcen ein tieferes Lernen fördern können.

Artikel 2 vergleicht die kognitive Belastung beim Lernen mit Video-Tutorials und im traditionellen Unterricht mithilfe von Herzratenvariabilitätsmessungen (HRV), insbesondere anhand des RMSSD-Indikators. Die Untersuchung mit denselben 45 Schüler:innen zeigt, dass Lernende während der Entwicklungsphase beim Einsatz von Video-Tutorials eine höhere kognitive Belastung erfuhren. Diese erhöhte Belastung korreliert mit intensiverer kognitiver Auseinandersetzung und tieferer Informationsverarbeitung.

Artikel 3 Artikel 3 analysiert die dynamische Beziehung zwischen Verhalten und Kognition der Lernenden anhand von Time-on-Task-Daten und HRV-Werten. Die Untersuchung wurde mit 45 Schüler:innen aus zwei berufsbildenden Schulen und einem Gymnasium durchgeführt. Die Ergebnisse zeigen eine signifikante negative Korrelation zwischen der aufgewendeten Lernzeit und der HRV. In Klassen mit Videoeinsatz zeigten sich stabilere kognitive Muster, während im traditionellen Unterricht stärkere Schwankungen beobachtet wurden.

Die zusammengeführten Ergebnisse der drei Studien verdeutlichen das Potenzial von Video-Tutorials im traditionellen Unterricht. Sie zeigen, dass Video-Tutorials die kognitive Belastung optimieren, die Lernbeteiligung fördern und potenziell die Lernergebnisse verbessern können. Dennoch bestehen einige Limitationen, wie etwa eine geringe Stichprobengröße, individuelle Unterschiede zwischen den Teilnehmenden und begrenzte Kontrolle über Störvariablen. Zukünftige Forschung sollte diese Einschränkungen gezielt adressieren und den Einsatz von Video-Tutorials weiter vertiefen.

Declaration of Candidate

Eidesstattliche Versicherung

Ich, Enqi Fan, versichere an Eides statt, dass ich die Dissertation mit dem Titel: „Video tutorials in traditional classrooms: The effects on cognitive load and time on task" selbst verfasst und keine anderen als die angegebenen Hilfsmittel benutzt habe.

Hamburg, 03.04.2025 Enqi Fan

Ort, Datum Unterschrift Doktorand:in

Erklärung

Darüber hinaus erkläre ich, dass ich keine kommerzielle Promotionsberatung in Anspruch genommen habe und die Arbeit nicht schon einmal in einem früheren Promotionsverfahren angenommen oder als ungenügend beurteilt worden ist.

Hamburg, 03.04.2025 Enqi Fan

Ort, Datum Unterschrift Doktorand:in

1. Introduction

1.1 Research Background

Video tutorials (VTs) are materials that combine audio and video information to provide step-by-step instruction to develop students' knowledge and skills (Noor et al., 2014; Ponzanelli et al., 2016). In recent years, technological developments and the widespread use of the Internet have made VTs an increasingly popular educational tool (Martin & Martin, 2015; Mayer et al., 2020). VTs have several advantages over traditional instruction. They can present complex concepts in a more straightforward way (Morain & Swarts, 2012) and convey knowledge in a more vivid form (Hong et al., 2016). In addition, video tutorials allow learners to control the pace of learning (de Koning et al., 2007; Martin, 2016) and enable them to manage their learning time more effectively through viewing strategies (Luke & Hogarth, 2011; Costley et al., 2020). Although many studies have shown that teaching with video tutorials has a positive impact on student efficiency, engagement, and learning outcomes (Lloyd & Robertson, 2011; Wells et al, 2012; Gonzalves et al. 2018; Hamas et al. 2019; Rizza et al. 2019; Rozi et al., 2020), there are also studies that suggest the opposite conclusion (Käfer et al., 2016; Ganier & de Vries, 2016; Zinn et al., 2021). Therefore, more in-depth research on the use of video tutorials is needed.

In order to understand the effectiveness of video tutorials, it is crucial to investigate the cognitive load of the learners. Cognitive load theory (CLT) provides an important theoretical framework for designing video tutorials (Noor et al. 2013). According to cognitive load theory, different types of cognitive resources are used in learning tasks, which include intrinsic load, extraneous load and germane load. These types of cognitive load affect the learning process and outcomes (Sweller, 1988; Sweller et al., 1998). Intrinsic cognitive load is determined by the complexity of the material and the learners' prior knowledge. Extraneous cognitive load is related to how the material is presented and organized. Germane cognitive load, on the other hand, is related to the construction of schemas (Sweller, 1988; Sweller et al. 1998; Leppink et al. 2013; Klepsch et al. 2017). Research indicates that video tutorials can affect cognitive load. For example, it may increase extraneous cognitive load and negatively affect learning (Mayer, 2005). However, insufficient cognitive load can also impede learning

(Leppink & van den Heuvel, 2015). Ideally, well-designed video tutorials should enhance germane cognitive load and thus facilitate learning while minimizing extraneous cognitive load (Sweller, 1994; Sweller et al., 2011). Although cognitive load is important in learning, few comparative studies between video tutorials and traditional instruction have explored types of cognitive load in depth over the past decade (Mutlu-Bayraktar et al., 2019). Therefore, research on different types of cognitive load is necessary to understand the effects of VT.

In addition, there is often a challenge in real classroom environments to understand whether students are truly engaged in learning, rather than just appearing to be attentive. Cognitive load theory suggests that when learners are less engaged, they use fewer cognitive resources in the learning process. Conversely, more engaged learners utilize more cognitive resources for deeper processing (Miller, 2015). Measuring students' cognitive load can help teachers understand how cognitively engaged their students are and determine what content is more likely to promote effective learning. However, most existing research focuses on measuring overall cognitive load (Costley & Lange, 2017), and almost none has examined changes in cognitive load at different phases of the learning process. Given that cognitive load varies throughout the learning process (Leppink et al., 2013; Paas et al., 2016), understanding the effectiveness of video tutorials at different phases of instruction can provide valuable insights for optimizing instructional strategies. Therefore, it is necessary to compare changes in cognitive load when using video tutorials versus traditional instruction in a real classroom setting.

Simultaneously, time-on-task can provide a new perspective on measuring student engagement and learning outcomes. Time-on-task is the amount of time students actively spend on learning tasks (Hesse, 1994). It is closely related to student engagement and learning outcomes (Buijs & Admiraal, 2013; Ghergulescu & Muntean, 2016; Spanjers et al., 2008) Numerous studies have shown a positive correlation between time-on-task and learning outcomes (Carvalho et al., 2017; Fisher, 1978; Fredrick et al., 1979; Kärner et al., 2015; Karweit & Slavin, 1982; Landers & Landers, 2014). Compared to non-standardized tests and homework, which are more subjectively influenced (Caspersen et al., 2017; Yorke, 2011), time-on-task can assess engagement and learning outcomes from a more objective perspective (Usart et al., 2013).

It is also worth considering that in a real classroom there are often many lesson phases of which the use of video tutorials is only one part, due to the design of the teacher. A classroom may contain an organization phase that is unrelated to the content, an introduction phase that is related to the topic, a development phase for learning new knowledge, a practice phase for applying the knowledge, and a debriefing phase in which the teacher and students discuss the learning results. Students may use video tutorials only during the development phase of learning new knowledge, while other instructional activities, such as organization and practice, may take place at other times. In order to better understand the impact of video tutorials in the classroom, it is necessary to investigate the different phases of the lesson, especially the development phase in which students acquire new knowledge and learn by using video tutorials.

In this dissertation, comparative studies of traditional classroom (TC) and video classroom (VC) were conducted to examine the effectiveness of video tutorials. Students' cognitive load on categorization was measured by a subjective scale, and changes in cognitive load were assessed by collecting biofeedback measures of heart rate variability (HRV). At the same time, students' time-on-task was coded through video and audio from the classroom as a way to assess student engagement at different phases of learning.

1.2 Theoretical Framework

1.2.1 Cognitive Load Theory (CLT)

In the late 1980s, John Sweller explained the interaction between long-term memory (Rumelhart, 1980) and limited working memory (Miller, 1956) based on schema theory (Chi et al., 1982; Larkin et al., 1980). He proposed the cognitive load theory, which suggests that human cognitive resources are limited. The process of learning and problem solving consumes cognitive resources and thus generates a certain amount of load. In early studies of cognitive load, cognitive load was divided into two categories: intrinsic cognitive load (related to schema construction) and extraneous cognitive load (not related to schema construction). As research on cognitive load deepened, researchers introduced germane cognitive load when they found that some of the cognitive loads produced effects that facilitated learning. The differences between the three types of cognitive load are as follows:

- **Intrinsic cognitive load:** Intrinsic cognitive load (ICL) is determined by the complexity of the learning material and the learner's prior knowledge (Sweller, 1988; Leppink et al., 2013; Klepsch et al., 2017). The complexity of learning materials is related to the element interactivity. Materials with low element interactivity have each element that can be learned independently of the others, so even though there are many elements, they do not require much working memory. Materials with high element interactivity cannot be learned independently, and multiple elements must be considered simultaneously in the learning process. Therefore, the higher the element interactivity, the higher intrinsic load (Sweller, 2010). Additionally, the learner's prior knowledge is also a factor that influences the intrinsic load. If the learner already has a richer prior knowledge of the domain covered by the learning content, the new knowledge can be more quickly categorized into existing schemas, thus reducing the load on working memory (Sweller et al., 2019). However, more working memory is required to process more of the learning content when the learner's prior knowledge is insufficient, resulting in an increased intrinsic load (Leppink et al., 2014).
- **Extraneous cognitive load:** Extraneous cognitive load (ECL) or ineffective cognitive load, which is related to how learning materials are presented and organized (Sweller et al., 2019). When learning materials are poorly designed, it

can cause learners to unnecessarily process elements that are not relevant to learning (Sweller et al., 1998). For example, learners may be asked to unnecessarily search within materials for information to solve a problem or for an unclear reference in an explanation (Paas et al., 2003). This can cause the learner to experience the split attention effect, resulting in an increase in extraneous load (Ayres & Sweller, 2005). Therefore, excessive extraneous load can interfere with learning and should be kept to a low level when designing instruction (Paas & Sweller, 2014).

- **Germane cognitive load:** Germane cognitive load (GCL) is considered as a load necessary for learning (Schnotz & Kürschner, 2007). More cognitive resources are allocated to intrinsic load when a learners' extraneous cognitive load is low. This results in the processing of elements from working memory and their transfer to long-term memory (schema construction). The cognitive resources used in this process are called germane cognitive load (Sweller et al., 1998; Leppink et al., 2013). Thus, germane load is a cognitive resource needed to deal with intrinsic cognitive load (Sweller et al., 2019). However, it is also due to the very close relationship between GCL and ICL that the concept of germane load has been controversial since it was proposed. Some researchers have argued that germane load is not independent of the other two cognitive loads, but rather uses the same theoretical foundations as intrinsic load, making it indistinguishable from intrinsic load (Kalyuga, 2011). Other researchers have argued that GCL is an active load and that high GCL is a cognitive resource that learners invest in, whereas ICL is a load that is passively experienced (Moreno & Park, 2010; Klepsch & Seufert, 2021). In our study, we aim to investigate whether video tutorials are effective in transforming information into schemas, and therefore we will use the three-factor cognitive load categorization.

Cognitive load and video tutorials

The aim of applying CLT in educational research is to reduce the cognitive load that hinders learning (Anmarkrud et al., 2019) and to promote the cognitive load that facilitates learning (Sweller et al., 1998). Thus, cognitive load is not just a by-product of the learning process but should be seen as a major determinant of the success of an instructional intervention (Paas et al., 2003; Kirschner, 2002). We hypothesize that effective video tutorials should be able to:

- Increase germane load by encouraging deep processing and schema construction (Sweller et al., 1998).
- Reduce extraneous load by designing instructional materials that are clear and concise (Sweller, 2010; Paas & Sweller, 2014).
- Enhance student engagement by mobilizing more cognitive resources while ensuring the above (Lan et al., 2019; Zheng et al., 2023).

1.2.2 Time-on-Task hypothesis

The time-on-task hypothesis was proposed by John B. Carroll in 1963. The hypothesis suggests that the more time students devote to learning, the more effective their learning will be (Carroll, 1963; Carroll, 1970). The longer the time-on-task, the more effective the learning task and instructional design assigned by the teacher (Gettinger & Seibert, 2002; Helmke, 2007; Lipowsky, 2006). Effective instructional strategies are able to increase students' time-on-task, which is important for achieving learning goals. This is because increasing time-on-task improves learning in all learning environments (Brown, 2001).

- **Engagement:** Time-on-task is an indicator of how engaged students are in learning activities (Buijs & Admiraal, 2013; Ghergulescu & Muntean, 2016; Spanjers et al. 2008)
- **Learning outcomes:** There is a positive correlation between time-on-task and learning outcomes (Carvalho et al.,2017; Fredrick et al, 1979; Fisher, 1978; Karweit & Slavin, 1982; Kärner et al., 2015; Landers & Landers, 2014).

Time-on-task and video tutorials

We hypothesize that effective video tutorials can increase time-on-task and thus improve learning engagement and learning outcomes by:

- Providing content that captures students' attention and makes them more focused on learning
- Allowing students to control their own learning pace, encouraging them to generate more effective learning time-on-task.

1.3 Research Methods

Participants

We conducted a seven-group controlled experiment in five vocational schools and one high school (two groups) in Germany. By having lessons from different types of schools and subjects, we were able to examine the effectiveness of the video tutorials in a variety of teaching and learning environments. This enhanced the generalizability of the results. The content of the lessons was based on the students' previous teaching schedules, ensuring that students had some knowledge of the subjects they were studying. Only students who participated in both classes were selected as samples to ensure that the observed cognitive load differences were not due to individual differences. To detect an effect size of Cohen's $d = .50$ with 90% power ($\alpha = .05$, two-tailed), G*Power 3.1 suggested that we needed 44 participants for a paired samples t-test (Faul et al., 2009; Serdar et al., 2021). The size of each class ranged from 10 to 20 students, and the six groups of classes ensured that the number of students participating in both lessons met the requirements.

Video and Audio Collection

We placed four HD cameras in the corners of the classroom to record instructional videos. At the same time, each student received a portable recorder to capture classroom audio.

Lesson phases and Time-on-task

We used the video analysis software Mangold interact (Mangold International GmbH, Arnstorf, Germany) to code the classroom videos. It can play multiple videos and audio simultaneously. The researchers coded the classroom videos according to the coding manual "*Classroom Observation Coding Manual System - Lesson Phases, Social Forms, and Time-on-Task*" (Fan & Siemon, 2024; see Appendix: coding manual). This coding system contains three coding manuals. The coding process was completed by multiple trained coders in the video lab.

Subjective Cognitive Load

We used Leppink et al.'s (2013) Cognitive Load Scale (CLS) to measure students' subjective cognitive load. The CLS consists of ten items grouped into three dimensions: ICL (three items, from 1 to 3), ECL (three items, from 4 to 6), and GCL (four items, from 7 to 10). The scale uses an 11-point Likert scale, where 0 means "not at all the case" and 10 means "completely the case." The reliability of the original version of the dimensions was Cronbach's α: ICL = .82, ECL = .75, and GCL = .82. In a meta-study examining the reliability of various cognitive load scales, Leppink et al.'s scale also achieved good results (Cronbach's α: ICL = .845, ECL = .759, GCL = .909) and has been widely used (Mutlu-Bayraktar et al., 2019; Krieglstein et al., 2022). Since the original version of the scale was used in a statistics course, we modified the statistics-related parts and had them translated into German by a native speaker. The scale results were analyzed using paired t-tests to find differences in cognitive load between classrooms.

Objective Cognitive Load

We used the Polar H10 heart rate monitor (Polar Electro Oy, Kempele, Finland) as the HRV collection device. This small, portable heart rate monitor has been shown to accurately measure RR intervals (Gilgen-Ammann et al., 2019; Speer et al., 2020; Moya-Ramon et al., 2022). The Polar H10 samples at 1000 Hz, and the data were transferred to iMotions software (9.3, iMotions, Copenhagen, Denmark) on four high-performance laptops via Bluetooth connection, with all data stored locally. Several methods can be used to analyze HRV, including time-domain analysis, frequency-domain analysis, and nonlinear dynamic analysis. Since we needed to calculate HRV for different phases of the lesson, with durations ranging from a few seconds to several minutes, we chose RMSSD from time-domain analysis as an indicator of cognitive load (Stein et al., 1994). RMSSD, which can be calculated even with short heart rate recordings, is the root mean square of successive differences between normal heartbeats and serves as an indicator of the parasympathetic nervous system (Thong et al., 2003; Pham et al., 2021). Therefore, a decrease in RMSSD values indicates higher cognitive load.

2. Overview of Individual Papers

2.1 Article 1: Video Tutorials in the Traditional Classroom: The Effects on Different Types of Cognitive Load

Published: June 2024, Technology, Knowledge and Learning

Research question:

- RQ 1. Are there differences in types of students' cognitive load between video classroom and traditional classroom?
- RQ 2. Are there correlations between the types of students' cognitive load in the video classroom and traditional classroom?

Research methods

Design: The study was conducted in two vocational schools and one high school (2 of the groups) in Germany. Each group was taught by four different teachers, each of whom used video tutorials in one lesson (video classroom: VC) and traditional teaching methods in another lesson (traditional classroom: TC). The content of each group is chosen by the instructor. The pace of learning strictly follows the original program of study. Only the development phase is replaced by a video tutorial in VC. Students were able to control the pace of the videos.

Participants: A total of 45 students (21 males, M_{age}= 19.42, SD_{age}= 2.54) participated in the study. The students were divided into four groups:

- Group A: Vocational school, topic "Light and Color", 15 students (3 males, M_{age}= 22.27, SD_{age}= 2.55).
- Group B: Vocational school, topic "Economy", 5 students (5 males, M_{age}= 18.40, SD_{age}= 0.55).
- Group C: High school, topic "Biology", 13 students (8 males, M_{age}= 18.00, SD_{age}= 0.41).
- Group D: High school, topic "Economics and Politics", 11 students (5 males, M_{age}= 17.83, SD_{age}= 0.83).

Research results

Differences in Cognitive Load:

Intrinsic Cognitive Load (ICL): The VC reported significantly lower ICL than the TC (VC: $M = 3.600$, $SD = 2.100$; TC: $M = 5.033$, $SD = 2.313$), $t = -4.507$, $p < .001$, $d = -0.672$.
Extraneous Cognitive Load (ECL): The VC reported slightly lower ECL than the TC, but the difference was not significant (VC: $M = 1.607$, $SD = 1.474$; TC: $M = 2.174$, $SD = 2.245$), $t = -1.688$, $p = .098$, $d = -0.252$.
Germane Cognitive Load (GCL): The VC reported significantly higher GCL than the TC (VC: $M = 7.444$, $SD = 1.681$; TC: $M = 6.200$, $SD = 1.786$), $t = 4.749$, $p < .001$, $d = 0.708$.

Correlations of Cognitive Load:

- ICL and ECL were positively correlated (VC: $r = 0.337$; TC: $r = 0.498$).
- ICL and GCL were negatively correlated (VC: $r = -0.442$; TC: $r = -0.451$).
- ECL and GCL were negatively correlated (VC: $r = -0.332$; TC: $r = -0.628$).

Meanwhile, to explore differences in correlations between different types of cognitive load in the two classrooms, we used Fisher's r-to-z transformation (Silver & Dunlap, 1987) to conduct a difference-in-difference analysis.

- Comparing ICL and ECL, the correlation for the video classroom is $z = 0.397$, while the traditional classroom is $z = 0.547$. The difference between the two classes is $z = -0.688$, which is not statistically significant, $p = 0.492$, $q = -0.150$.
- Comparing ICL and GCL, the correlation for the video classroom is $z = -0.475$ compared to $z = -0.486$ for the traditional classroom. The observed difference between the two classes is $z = 0.052$, which is not statistically significant, $p = 0.959$, $q = 0.011$.
- Comparing ECL and GCL, the correlation for the video classroom is $z = -0.345$, the correlation for the traditional classroom is $z = -0.738$. The difference between these correlations is $z = 1.801$, $p = 0.072$, $q = 0.393$. This suggests a trend towards statistical significance.

These results suggest that the use of video tutorials can be effective in reducing students' intrinsic cognitive load and increasing germane cognitive load, thereby facilitating deeper learning. Furthermore, the correlation between different types of cognitive load suggests that students dynamically adapt the use of cognitive resources to different tasks and content during the learning process.

2.2 Article 2: Comparing cognitive load during video versus traditional classroom instruction based on heart-rate variability measures

Submitted: July 2024, Computers & Education

Research question:

- RQ3: Do students who use video tutorial compared to traditional instruction have a difference in cognitive load during the development phase of learning?

Research methods

Design: The study was conducted in two vocational schools and one high school (2 of the groups) in Germany. Each group was taught by four different teachers, each of whom used video tutorials in one lesson (video classroom: VC) and traditional teaching methods in another lesson (traditional classroom: TC). The content of each group is chosen by the instructor. The pace of learning strictly follows the original program of study. Only the development phase is replaced by a video tutorial in VC. Students were able to control the pace of the videos.

Measurement Tool:
Heart rate variability (HRV) data were collected using a Polar H10 heart rate belt. RMSSD (Root Mean Square Deviation) was used as an indicator to assess cognitive load.

Participants: A total of 45 students (21 males, M_{age}= 19.87, SD_{age}= 3.66) participated in the study. The students were divided into four groups:

- Group A: Vocational school, topic "Unemployment", 10 students (1 male, M_{age}= 23.00, SD_{age}= 3.94).
- Group B: Vocational school, topic "Security Management", 10 students (7 males, M_{age}= 21.60, SD_{age}= 4.90).
- Group C: High school, topic "Biology", 13 students (8 males, M_{age}= 18.00, SD_{age}= 0.41).
- Group D: High school, topic "Economics and Politics", 12 students (5 males, M_{age}= 17.83, SD_{age}= 0.83).

Research results

Overall Finding:

Paired-samples t-tests revealed that students' cognitive load in the development phase was significantly higher in VC (RMSSD percent change: -4.213, standard deviation: 8.808) than in TC (RMSSD percent change: 2.512, standard deviation: 7.183), $t(44) = -4.205$, $p < .001$, Cohen's $d = -.627$.

Comparison of each group:

- Group A: Cognitive load was slightly higher in the VC development phase but did not reach statistical significance. mean RMSSD percent change for VC = -5.979%, mean RMSSD percent change for TC = -3.621%, $t(9) = -.712$, $p = 0.495$, Cohen's $d = -.225$.
- Group B: Cognitive load was higher in the VC development phase and approached statistical significance. mean RMSSD percent change for VC = -6.328%, mean RMSSD percent change for TC = 2.669%, $t(9) = -1.934$, $p = 0.085$, Cohen's $d = -.612$.
- Group C: Cognitive load was significantly higher in the TC development phase. mean RMSSD percent change in VC = 0.038%, mean RMSSD percent change in TC = 7.059%, $t(12) = -3.460$, $p = 0.005$, Cohen's $d = -.960$.
- Group D: Cognitive load was higher in VC development phase and reached statistical significance. mean RMSSD percentage change in VC = -5.586%, mean RMSSD percentage change in TC = 2.566%, $t(11) = -2.683$, $p = 0.021$, Cohen's $d = -.775$.

The findings suggest that during the development phase, students invest more cognitive resources in VC than TC. High cognitive load is usually accompanied by high engagement, which indicates that students are more actively mobilizing cognitive resources for deeper learning processing when using video tutorials. Based on the results of the previous study, we can assume that this high cognitive load is not a negative, but rather a positive and favorable cognitive load that helps students transform knowledge into long-term memory.

2.3 Article 3: From Heartbeats to Actions: Multimodal Learning Analytics of Cognitive and Behavior Engagement in Real Classrooms

Submitted: December 2024, Learning and Instruction

Research question:

- RQ4: Is there a difference in HRV changes for different types of tasks in the video learning classroom and the traditional learning classroom?
- RQ5: Is there a correlation between HRV changes and Time-on-task changes for different types of tasks in the video learning classroom and the traditional learning classroom?

Research methods

Design: The study was conducted in three vocational schools and one high school in Germany. Each group was taught by four different teachers, each of whom used video tutorials in one lesson (video learning classroom: VLC) and traditional teaching methods in another lesson (traditional learning classroom: TLC). The content of each group is chosen by the instructor. The pace of learning strictly follows the original program of study. Only the development phase is replaced by a video tutorial in VLC. Students were able to control the pace of the videos.

Participants: A total of 45 students (21 males, M_{age}= 19.75, SD_{age}= 3.67) participated in the study. The students were divided into four groups:

- Group A: Vocational school, topic “Security Management”, 10 students (7 males, M_{age}= 21.60, SD_{age}= 4.90).
- Group B: Vocational school, topic “Unemployment”, 9 students (1 male, M_{age}= 23.00, SD_{age}= 3.94).
- Group C: High school, topic “Biology”, 14 students (8 males, M_{age}= 18.00, SD_{age}= 0.41).
- Group D: High school, topic “Economics and Politics”, 12 students (5 males, M_{age}= 17.83, SD_{age}= 0.83).

Data analysis: This study coded the time-on-task of each participant and generated raw data with a code every 10 seconds (Time-on-task-10s). We then calculated the percentage of time spent on the task per minute (Time-on-task%-60s). We also calculated the RMSSD every 10 seconds (RMSSD-10s) and every 60 seconds (RMSSD-60s) for each student to synchronize the HRV data with the Time-on-task-10s and Time-on-task%-60s data.

Research results

- The Kruskal-Wallis test results showed that there were differences between the **RMSSD%-10s** in the **Time-on-task-10s** codes of VLC and TLC.
- Dunn's test results showed that Code 3 (real learning time) occupied the main proportion in both VLC and TCL.
- Spearman correlation analysis results show that there is a significant negative correlation between **Time-on-task%-60s** and **RMSSD%-60s** in both classroom environments, but the strength of the correlation is different. In VLC, the correlation coefficient was weaker (ρ = -0.1621, p = 0.014). In contrast, the correlation coefficient in TLC is stronger (ρ = -0.2184, p < 0.001), indicating that students' RMSSD%-60s fluctuations are more closely related to changes in On-Task.

The study reveals the characteristics of students' learning states in different classroom environments by integrating behavioral and biofeedback data. The results show that time-on-task was negatively correlated with HRV, and combining the two provided a more comprehensive understanding of students' learning states. In the VLC, students' cognitive states were relatively stable, especially during the development phase. However, in the TLC, students' cognitive levels fluctuated more. This indicates that video tutorials play an important role in mobilizing students' cognitive resources and reducing fluctuations in cognitive load.

3. Discussion and Conclusion

3.1 Research Implications

This dissertation analyzed the effects of video tutorials on students' cognitive load and time-on-task at different stages of instruction. The results demonstrate the potential benefits of video tutorials in instruction. Therefore, we propose the following recommendations:

a) **Researchers should promote classroom empirical research from multiple perspectives**

The studies were conducted in real classroom settings, which verified the feasibility of empirical research on cognition and behavior. By analyzing individual lesson phases, the importance of the timing of teaching interventions in the classroom was highlighted. In the future, researchers should focus not only on the video tutorials themselves, but also on their implementation strategies.

b) **Teachers should use video tutorials more often in class**

The studies have shown that the use of video tutorials is effective not only in online teaching but also in traditional classrooms. Video tutorials significantly improve students' germane cognitive load and time-on-task. Therefore, teachers should consider using video tutorials more often in the class.

c) **Teachers should consider the timing of using video tutorials in instructional design**

The studies indicate that replacing traditional teaching with video tutorials during developmental phases is particularly effective. At this phase, students are in the process of initial understanding and construction of new concepts and skills. Therefore, in instructional design, teachers should appropriately integrate video tutorials into the developmental phase based on the teaching content and learning progress.

d) **Teacher training should develop the ability to use video tutorials in teaching**

To ensure that video tutorials are used effectively in the teaching process, teachers need to develop specific skills. Therefore, both pre-service and in-service teacher training should include content on how to use video tutorials effectively. This includes how to select video tutorials based on cognitive load theory, how to properly integrate video tutorials into teaching, and emergency strategies for technical issues.

e) **School administrators should provide comprehensive technical support for the use of video tutorials**

The effective use of video tutorials not only relies on teachers' individual skills but also on the support of school administrators in terms of technology and resources. Thus, school administrators should provide teachers with comprehensive technical support and resource allocation whenever possible. This includes the updating and maintenance of hardware equipment, support for software platforms, as well as technical training and services. Such a support system ensures that teachers can smoothly apply video tutorials during teaching, thereby enabling technology to truly serve teaching.

In conclusion, video tutorials should be seen as an important component of effective classroom teaching. Their use requires thoughtful design of the teaching, proper training of teachers, and ongoing support from school administrators. By focusing on these areas, video tutorials can provide a richer and more engaging learning experience for students.

3.2 Limitations and Future Research Directions

Sample Size: Although each study achieved Cohen's d = .50 with 90% power (α = .05, two-tailed) effect size, the sample size was still small. Small sample sizes may lead to increased randomization of results. The sample size should be expanded in future studies to ensure that the findings can be generalized to a wider group of students.

Individual differences: Participants came from a variety of schools and backgrounds and there were large differences in learning styles and motivation between individuals. This was particularly obvious in the results for time-on-task. Analysis of individual difference factors such as students' learning styles and motivation could be added to future research and the influence of these factors on the effectiveness of the video tutorials could be discussed.

Cognitive load measurement: Although both subjective and objective methods were used to measure cognitive load in the study, there are still some limitations. First, this study was conducted on the assumption that germane load exists. However, the theoretical definition of GCL is still controversial. Some researchers have pointed out that the boundary between GCL and ICL is difficult to distinguish clearly. This shows that CLT still needs further theoretical clarification and empirical research support. Second, although many studies have verified the effectiveness of HRV in measuring cognitive load, this is not absolute. As a biofeedback indicator, HRV can be affected by external environment or individual state. Future studies should consider introducing multiple biofeedback or behavioral indicators for cross-validation.

Control variables: Although variables such as teacher, pace of instruction, etc. were controlled as much as possible in the study, some other variables may still have an impact on the results. For example, the design and quality of the video tutorials may have an effect on the results of the study, but these factors were not fully controlled in the current study. Future research should further standardize the experimental conditions and strictly control variables such as teachers' teaching style, course content and video tutorial quality to ensure the reliability and validity of the results.

Long-term effects: The studies are mainly short-term experiments and lacks an exploration of the long-term effects of video tutorials. The fact that learning performance in the short-term may not be representative of students' long-term learning effects limits the applicability of the study's findings. In future studies, long-term follow-up studies should be designed to regularly assess students' ability to apply their knowledge. This would allow exploration of the effects of video tutorials on students long-term learning outcomes for more comprehensive conclusions.

4. Reference

Anmarkrud, Ø., Andresen, A., & Bråten, I. (2019). Cognitive Load and Working Memory in Multimedia Learning: Conceptual and Measurement Issues. *Educational Psychologist*, *54*(2), 61–83. https://doi.org/10.1080/00461520.2018.1554484

Brown, K. G. (2001). Using computers to deliver training: Which employees learn and why? *Personnel psychology, 54*(2), 271-296.

Buijs, M., & Admiraal, W. (2013). Homework assignments to enhance student engagement in secondary education. *European journal of psychology of education*, 28, 767-779.

Carroll, J. B. (1963). A model of school learning. *Teachers college record, 64*(8), 1-9.

Carroll, J. B. (1970). Problems of measurement related to the concept of learning for mastery. *Educational Horizons, 48*(3), 71-80.

Carvalho PF, McLaughlin EA, Koedinger K. (2017). Is there an explicit learning bias? *Students beliefs, behaviors and learning outcomes.* InCogSci.

Caspersen, J., Smeby, J. C., & Olaf Aamodt, P. (2017). Measuring learning outcomes. *European Journal of Education, 52*(1), 20-30.

Chi, M., Glaser, R., & Rees, E. (1982). Expertise in problem solving. In R. Sternberg (Ed.), *Advances in the psychology of human intelligence* (pp. 7–75). Hillsdale, NJ: Erlbaum.

Costley, J., Fanguy, M., Lange, C., & Baldwin, M. (2020). The effects of video lecture viewing strategies on cognitive load. *Journal of Computing in Higher Education*, *33*(1), 19–38. https://doi.org/10.1007/s12528-020-09254-y

de Koning, B. B., Tabbers, H. K., Rikers, R. M., & Paas, F. (2007). Attention cueing as a means to enhance learning from an animation. *Applied Cognitive Psychology*, *21*(6), 731–746. https://doi.org/10.1002/acp.1346

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. https://doi.org/10.3758/brm.41.4.1149

Fisher, C. W. (1978). Teaching Behaviors, Academic Learning Time and Student Achievement: Final Report of Phase III-B, *Beginning Teachers Evaluation Study.* Technical Report V-1

Fredrick, W. C., Walberg, H. J., & Rasher, S. P. (1979). Time, Teacher Comments, and Achievement in Urban High Schools. *The Journal of Educational Research, 73*(2), 63–65. https://doi.org/10.1080/00220671.1979.10885209

Ganier, F., & de Vries, P. (2016). Are instructions in video format always better than photographs when learning manual techniques? The case of learning how to do sutures. *Learning and Instruction*, *44*, 87–96. https://doi.org/10.1016/j.learninstruc.2016.03.004

Gettinger, M., & Seibert, J. K. (2002). *Best practices in increasing academic learning time.*

Ghergulescu, I., & Muntean, C. H. (2016). ToTCompute: A novel EEG-based TimeOnTask threshold computation mechanism for engagement modelling and monitoring. *International Journal of Artificial Intelligence in Education*, 26, 821-854.

Gilgen-Ammann, R., Schweizer, T., & Wyss, T. (2019). RR interval signal quality of a heart rate monitor and an ECG Holter at rest and during exercise. *European Journal of Applied Physiology*, *119*(7), 1525–1532. https://doi.org/10.1007/s00421-019-04142-5

Gonzalves, A., Verhaeghe, C., Bouet, P. E., Gillard, P., Descamps, P., & Legendre, G. (2018). Effect of the use of a video tutorial in addition to simulation in learning the maneuvers for shoulder dystocia. *Journal of Gynecology Obstetrics and Human Reproduction*, *47*(4), 151–155. https://doi.org/10.1016/j.jogoh.2018.01.004

Hamas, A., Maryono, D., & Efendi, A. (2019). The Effectiveness Of The Use Of Blended Learning Model Based On Video Tutorial Reviewed From Students' Understanding In The Basic Graphic Learning In Banyudono State Vocational School. *Journal of Informatics and Vocational Education*, *2*(2). https://doi.org/10.20961/joive.v2i2.37972

Helmke, A. (2007). Aktive Lernzeit optimieren. Was wissen wir über effiziente Klassenführung? *In Pädagogik (Weinheim), 59*(5), 44-49.

Hesse, H. G. (1994). *Lehr-Lern-Zeit und Lernerfolg aus psychologischer Sicht* (pp. 143-161).

Hong, J., Pi, Z., & Yang, J. (2016). Learning declarative and procedural knowledge via video lectures: Cognitive Load and learning effectiveness. *Innovations in Education and Teaching International*, *55*(1), 74–81. https://doi.org/10.1080/14703297.2016.1237371

Käfer, V., Kulesz, D., & Wagner, S. (2016). *What Is the Best Way for Developers to Learn New Software Tools? A Small Empirical Comparison between a Text and a Video Tutorial*. https://doi.org/10.7287/peerj.preprints.2413v2

Kalyuga, S. (2011). Cognitive Load Theory: How Many Types of Load Does It Really Need? *Educational Psychology Review*, *23*(1), 1–19. https://doi.org/10.1007/s10648-010-9150-7

Kärner, T., Egloffstein, M., Binöder, F., Frötschl, C., & Schley, T. (2015). *Workload, time-on-task, and learning outcome in online learning for beginning students*.

Karweit, N., & Slavin, R. E. (1982). Time-on-task: Issues of timing, sampling, and definition. *Journal of educational psychology, 74*(6), 844.

Kirschner, P. A. (2002). Cognitive load theory: implications of cognitive load theory on the design of learning. *Learning and Instruction*, *12*(1), 1–10. https://doi.org/10.1016/s0959-4752(01)00014-7

Klepsch, M., & Seufert, T. (2021). Making an Effort Versus Experiencing Load. *Frontiers in Education*, *6*. https://doi.org/10.3389/feduc.2021.645284

Klepsch, M., Schmitz, F., & Seufert, T. (2017). Development and Validation of Two Instruments Measuring Intrinsic, Extraneous, and Germane Cognitive Load. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.01997

Krieglstein, F., Beege, M., Rey, G. D., Ginns, P., Krell, M., & Schneider, S. (2022). A Systematic Meta-analysis of the Reliability and Validity of Subjective Cognitive Load Questionnaires in Experimental Multimedia Learning Research. *Educational Psychology Review*, *34*(4), 2485–2541. https://doi.org/10.1007/s10648-022-09683-4

Lan, C.-H., Sheng, M.-H., Hsu, Y.-C., & Shiue, Y.-M. (2019). Influence of online and face-to-face collaboration and learning style on cognitive load and engagement in a health introductory course. *Proceedings of the Third International Conference on Medical and Health Informatics 2019 - ICMHI 2019*. https://doi.org/10.1145/3340037.3340069

Landers, R. N., & Landers, A. K. (2014). An empirical test of the theory of gamified learning: The effect of leaderboards on time-on-task and academic performance. *Simulation & Gaming, 45*(6), 769-785.

Larkin, J. H., McDermott, J., Simon, D. P., & Simon, H. A. (1980). Models of Competence in Solving Physics Problems*. *Cognitive Science*, *4*(4), 317–345. https://doi.org/10.1207/s15516709cog0404_1

Leppink, J., & Van den Heuvel, A. (2015). The evolution of cognitive load theory and its application to medical education. *Perspectives on Medical Education*, *4*(3), 119–127. https://doi.org/10.1007/s40037-015-0192-x

Leppink, J., Paas, F., Van der Vleuten, C. P., Van Gog, T., & Van Merriënboer, J. J. (2013). Development of an instrument for measuring different types of cognitive load. *Behavior Research Methods*, *45*(4), 1058–1072. https://doi.org/10.3758/s13428-013-0334-1

Leppink, J., Paas, F., Van der Vleuten, C. P., Van Gog, T., & Van Merriënboer, J. J. (2013). Development of an instrument for measuring different types of cognitive load. *Behavior Research Methods*, *45*(4), 1058–1072. https://doi.org/10.3758/s13428-013-0334-1

Leppink, J., Paas, F., van Gog, T., van der Vleuten, C. P. M., & van Merriënboer, J. J. G. (2014). Effects of pairs of problems and examples on task performance and different types of cognitive load. *Learning and Instruction*, *30*, 32–42. https://doi.org/10.1016/j.learninstruc.2013.12.001

Lipowsky, F. (2006). Auf den Lehrer kommt es an. Empirische Evidenzen für Zusammenhänge zwischen Lehrerkompetenzen, *Lehrerhandeln und dem Lernen der Schüler* (pp. 47-70).

Lloyd, S. A., & Robertson, C. L. (2011). Screencast Tutorials Enhance Student Learning of Statistics. *Teaching of Psychology*, *39*(1), 67–71. https://doi.org/10.1177/0098628311430640

Luke, B., & Hogarth, K. (2011). Developing and enhancing independent learning skills. *Accounting Research Journal*, *24*(3), 290–310. https://doi.org/10.1108/10309611111187019

Martin, N. A., & Martin, R. (2015). Would You Watch It? Creating Effective and Engaging Video Tutorials. *Journal of Library & Information Services in Distance Learning*, *9*(1–2), 40–56. https://doi.org/10.1080/1533290x.2014.946345

Martin, P. A. (2016). Tutorial video use by senior undergraduate electrical engineering students. *Australasian Journal of Engineering Education*, *21*(1), 39–47. https://doi.org/10.1080/22054952.2016.1259027

Mayer, R. E. (2005). Cognitive Theory of Multimedia Learning. *The Cambridge Handbook of Multimedia Learning*, 31–48. https://doi.org/10.1017/cbo9780511816819.004

Miller, B. W. (2015). Using reading times and eye-movements to measure cognitive engagement. *Educational Psychologist*, 50(1), 31–42. doi:10.1080/00461520.2015.100406

Morain, M., & Swarts, J. (2012). YouTutorial: A Framework for Assessing Instructional Online Video. *Technical Communication Quarterly*, *21*(1), 6–24. https://doi.org/10.1080/10572252.2012.626690

Moreno, R., & Park, B. (2010). Cognitive Load Theory: Historical Development and Relation to Other Theories. *Cognitive Load Theory*, 9–28. https://doi.org/10.1017/cbo9780511844744.003

Moya-Ramon, M., Mateo-March, M., Peña-González, I., Zabala, M., & Javaloyes, A. (2022). Validity and reliability of different smartphones applications to measure HRV during short and ultra-short measurements in elite athletes. *Computer Methods and Programs in Biomedicine*, *217*, 106696. https://doi.org/10.1016/j.cmpb.2022.106696

Mutlu-Bayraktar, D., Cosgun, V., & Altan, T. (2019). Cognitive load in multimedia learning environments: A systematic review. *Computers & Education*, *141*, 103618. https://doi.org/10.1016/j.compedu.2019.103618

Noor, N. M., Aini, M., & Hamizan, N. I. (2014). Video based learning embedded with cognitive load theory: Visual, auditory, and kinaesthetic learners' Perspectives. *2014 International Conference on Teaching and Learning in Computing and Engineering*. https://doi.org/10.1109/latice.2014.19

Noor, N. M., Hamizan, N. I., & Rahim, R. A. (2013). The framework for learning using video based on cognitive load theory among visual learners. *2013 IEEE 5th Conference on Engineering Education (ICEED)*. https://doi.org/10.1109/iceed.2013.6908295

Paas, F., & Sweller, J. (2014). Implications of Cognitive Load Theory for Multimedia Learning. *The Cambridge Handbook of Multimedia Learning*, 27–42. https://doi.org/10.1017/cbo9781139547369.004

Paas, F., Renkl, A., & Sweller, J. (Eds.). (2016). *Cognitive load theory: A Special issue of educational psychologist*. Routledge.

Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. (2003). Cognitive Load Measurement as a Means to Advance Cognitive Load Theory. *Educational Psychologist*, *38*(1), 63–71. https://doi.org/10.1207/s15326985ep3801_8

Pham, T., Lau, Z. J., Chen, S. H., & Makowski, D. (2021). Heart rate variability in psychology: A review of HRV Indices and an analysis tutorial. *Sensors*, *21*(12), 3998. https://doi.org/10.3390/s21123998

Ponzanelli, L., Bavota, G., Mocci, A., Di Penta, M., Oliveto, R., Hasan, M., Russo, B., Haiduc, S., & Lanza, M. (2016). Too long; didn't watch! extracting relevant fragments from software development video tutorials. *Proceedings of the 38th International Conference on Software Engineering*. https://doi.org/10.1145/2884781.2884824

Rizza, M., Widho Wati, C., & Dardiri, A. (2019). Transforming Digital Learning in Vocational High School 21st Century. *Proceedings of the 2nd International Conference on Vocational Education and Training (ICOVET 2018)*. https://doi.org/10.2991/icovet-18.2019.65

Rozi, F., Abdul Hamid, K., & Panjaitan, K. (2020). Development of Tutorial Video Media Based on Project Based Learning in Class XI State Vocational School 1 Pakam Lubuk. *Journal of Physics: Conference Series*, *1485*(1), 012056. https://doi.org/10.1088/1742-6596/1485/1/012056

Rumelhart, D. E. (1980). On Evaluating Story Grammars*. *Cognitive Science*, *4*(3), 313–316. https://doi.org/10.1207/s15516709cog0403_5

Schnotz, W., & Kürschner, C. (2007). A Reconsideration of Cognitive Load Theory. *Educational Psychology Review*, *19*(4), 469–508. https://doi.org/10.1007/s10648-007-9053-4

Serdar, C. C., Cihan, M., Yücel, D., & Serdar, M. A. (2021). Sample size, power and effect size revisited: simplified and practical approaches in pre-clinical, clinical and laboratory studies. *Biochemia Medica*, *31*(1), 27–53. https://doi.org/10.11613/bm.2021.010502

Spanjers, D. M., Burns, M. K., & Wagner, A. R. (2008). Systematic direct observation of time on task as a measure of student engagement. *Assessment for Effective Intervention*, 33(2), 120-126.

Speer, K. E., Semple, S., Naumovski, N., & McKune, A. J. (2020). Measuring heart rate variability using commercially available devices in healthy children: A validity and Reliability Study. *European Journal of Investigation in Health, Psychology and Education*, *10*(1), 390–404. https://doi.org/10.3390/ejihpe10010029

Stein, P. K., Bosner, M. S., Kleiger, R. E., & Conger, B. M. (1994). Heart rate variability: A measure of cardiac autonomic tone. *American Heart Journal*, *127*(5), 1376–1381. https://doi.org/10.1016/0002-8703(94)90059-0

Sweller, J. (1988). Cognitive Load During Problem Solving: Effects on Learning. *Cognitive Science*, *12*(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*, *4*(4), 295–312. https://doi.org/10.1016/0959-4752(94)90003-5

Sweller, J. (2010). Element Interactivity and Intrinsic, Extraneous, and Germane Cognitive Load. *Educational Psychology Review*, *22*(2), 123–138. https://doi.org/10.1007/s10648-010-9128-5

Sweller, J., Ayres, P., & Kalyuga, S. (2011). Measuring Cognitive Load. *Cognitive Load Theory*, 71–85. https://doi.org/10.1007/978-1-4419-8126-4_6

Sweller, J., van Merriënboer, J. J., & Paas, F. (2019). Cognitive Architecture and Instructional Design: 20 Years Later. *Educational Psychology Review*, *31*(2), 261–292. https://doi.org/10.1007/s10648-019-09465-5

Sweller, J., van Merrienboer, J. J., & Paas, F. G. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, *10*(3), 251–296. https://doi.org/10.1023/a:1022193728205

Thong, T., Li, K., McNames, J., Aboy, M., & Goldstein, B. (2003). Accuracy of ultra-short heart rate variability measures. *Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEEE Cat. No.03CH37439)*. https://doi.org/10.1109/iembs.2003.1280405

Usart, M., Romero, M., & Barberà, E. (2013). Measuring students' Time Perspective and Time on Task in GBL activities. *ELearn Center Research Paper Series, 40-51.*

Wells, J., Barry, R. M., & Spence, A. (2012). Using Video Tutorials as a Carrot-and-Stick Approach to Learning. *IEEE Transactions on Education*, *55*(4), 453–458. https://doi.org/10.1109/te.2012.2187451

Yorke, M. (2011). Summative assessment: dealing with the 'measurement fallacy'. *Studies in Higher Education, 36*(3), 251-273.

Zheng, X., Ma, Y., Yue, T., & Yang, X. (2023). Effects of different types of cues and self-explanation prompts in instructional videos on Deep learning: Evidence from multiple data analysis. *Educational Technology Research and Development*, *71*(3), 807–831. https://doi.org/10.1007/s11423-023-10188-2

Zinn, B., Tenberg, R., & Pittich, D. (2021). Erklärvideos – im naturwissenschaftlich-technischen Unterricht eine Alternative zu Texten? [explanatory videos as an alternative to text-based learning using the problem-solving methodology SPALTEN as an example?] *Journal of Technical Education*, 9(2), 168-187. https://doi.org/10.48513/joted.v9i2.217

Appendix：Publications

Article1：Video Tutorials in the Traditional Classroom: The Effects on Different Types of Cognitive Load

Article 2：Comparing cognitive load during video versus traditional classroom instruction based on heart-rate variability measures

Article 3：From Heartbeats to Actions: Multimodal Learning Analytics of Cognitive and Behavior Engagement in Real Classrooms

Coding Manual：Classroom Observation Coding Manual System - Lesson Phases, Social Forms, and Time-on-Task

Technology, Knowledge and Learning
https://doi.org/10.1007/s10758-024-09754-1

ORIGINAL RESEARCH

Video Tutorials in the Traditional Classroom: The Effects on Different Types of Cognitive Load

Enqi Fan[1] · **Matt Bower**[2] · **Jens Siemon**[1]

Accepted: 18 June 2024

Abstract

Are video tutorials better teachers? This pilot study examined the effects of video tutorials on different types of cognitive load. Participating students (N=45) attended two classrooms: a video tutorial-based classroom, and a traditional instruction-based classroom. The cognitive load scales indicated differences in cognitive load between the video classroom and the traditional classroom. Video tutorials decreased students' intrinsic load (t = -4.507, p<.001, d=−0.672) and increased germane load (t=4.749, p<.001, d=0.708) but did not affect extraneous load (t = -1.688, p=.098, d=−0.252). The results also indicated additivity for different types of cognitive load in the two classrooms. In general, our results demonstrate that video tutorials are a promising form of instructional material, especially to facilitate more effective and deeper learning.

Keywords Video tutorial · Video-based learning · Cognitive load

1 Introduction

With the development of technology and the accessibility of the Internet, video tutorials (VTs) have become an increasingly popular teaching tool in recent years (Martin & Martin, 2015; Mayer et al., 2020). Video tutorials have many advantages over traditional instructional materials: they can present complex concepts in a more intuitive way (Morain & Swarts, 2012), learners can control the pace of learning while using the video tutorials (de Koning et al., 2007; Martin, 2016), and through repeated viewing of the video tutorial, learners can manage and organize their study time more efficiently (Luke & Hogarth, 2011). Although many studies indicate that teaching with video tutorials has a positive effect on students' efficiency, engagement and learning outcomes (Lloyd & Robertson, 2011; Wells

✉ Enqi Fan
enqi.fan@uni-hamburg.de

[1] Department of Vocational and Business Education, University of Hamburg, Sedanstraße 19, 22119 Hamburg, Germany

[2] School of Education, Macquarie University, Building 29WW, Balaclava Rd North Ryde, NSW Sydney 2109, Australia

et al., 2012; Van der Meij & Van der Meij, 2014b; Martin, 2016; van der Meij & van der Meij, 2016; Wahyudi et al., 2017; Gonzalves et al., 2018; Hamas et al., 2019; Rizza et al., 2019; Rozi et al., 2020), there are also studies that show the opposite (Käfer et al., 2016; Ganier & de Vries, 2016; Zinn et al., 2021). This means that more research should focus on the conditions under which video tutorials improve learning outcomes and how to optimize them (Van der Meij & Van der Meij, 2014a).

In order to understand the effectiveness of video tutorials in-depth, it is necessary to examine learners' cognitive load. Cognitive Load Theory (CLT) provides an important theoretical foundation that guides the design of video tutorials (Noor et al., 2013). CLT identifies the different types of cognitive resources that individuals utilize in the course of completing a learning task. The three types of cognitive load utilized – intrinsic, extraneous, and germane – affect the task process and hence the task outcome (Sweller, 1988; Sweller et al., 1998). Intrinsic cognitive load is determined by the complexity of the material and the prior knowledge of the learner. Extraneous cognitive load is related to the way the learning material is presented and organized. Germane cognitive load is related to schema construction (Sweller, 1988; Sweller et al., 1998; Leppink et al., 2013; Klepsch et al., 2017).

Research suggests that video tutorials can have an impact on cognitive load (Paas et al., 2008). For example, it can lead to an increase in extraneous cognitive load and thus negatively affect learning (Mayer, 2005). But low cognitive load is also not beneficial for learning (Leppink & van den Heuvel, 2015). Ideally, a good video tutorial should allow learners to increase germane load that enhances learning and decreases extraneous load that is harmful to learning (Sweller, 1994; Sweller et al., 2011). However, in the past decade, there has been little research analyzing the sorts of cognitive load that manifests while learning through video tutorials. In addition, the types of cognitive load have rarely been explored in depth in comparative studies of video tutorials and traditional teaching methods. This make it difficult to link the findings of these comparative studies to the theoretical foundations (Mutlu-Bayraktar et al., 2019). Therefore, in the current research, there is a lack of research that compares the different types of cognitive load when video tutorials are being used in real classroom settings. The study aims to fill this gap by examining the differences in the types of cognitive load that arise during video tutorial-based classes (VC) and traditional instruction-based classes without video support (TC), in order to better understanding the different cognitive impacts of the two approaches.

2 Research Question

RQ (1) Are there differences in types of students' cognitive load between video classroom (VC) and traditional classroom (TC)?

RQ (2) Are there correlations between the types of students' cognitive load in the video classroom (VC) and traditional classroom (TC)?

3 The Literature Review

3.1 Cognitive Load Theory

Cognitive scientists have established that working memory or short-term memory is limited (Miller, 1956; Broadbent, 1958; Brown, 1958; Peterson & Peterson, 1959). Therefore, if the amount of information provided during instruction exceeds the learner's short-term memory capacity, the extra information is useless (Kalyuga & Sweller, 2014). In contrast, long-term memory (LTM) has a virtually unlimited storage capacity and provides a more permanent repository of knowledge and abilities (Bower, 2014). Memory research has shown that learners' prior knowledge experiences affect their recall of learning material (Bartlett, 1932), and these experiences are stored in long-term memory as schemas (Rumelhart, 1980). Once a schema is developed, it remains stable over long periods of time, allowing people to encode and classify information that has been or will be acquired. These cognitive processes are automatic and do not require conscious control or resource consumption (Rumelhart, 1980; Kirschner, 2002; Paas et al., 2003b).

In 1988, John Sweller explained the interaction between limited working memory (Miller, 1956) and long-term memory based on schema theory (Chi et al., 1982; Larkin et al., 1980). He provided a more comprehensive and systematic discussion of CLT from the perspective of cognitive resource allocation. CLT emphasizes that human cognitive resources are limited. The process of learning and problem solving consumes cognitive resources and thus generates a certain load. The purpose of applying CLT in the classroom is to reduce the cognitive load that hinders learning (Anmarkrud et al., 2019), and to promote the cognitive load that favors learning (Sweller et al., 1998). Therefore, cognitive load is not just a by-product of the learning process but should be considered a major determinant of the success of instructional interventions (Paas et al., 2003b; Kirschner, 2002). Instruction must consider how to avoid cognitive overload for the learner in a limited amount of time and be able to store knowledge in long-term memory.

3.2 Types of Cognitive Load

The discourse surrounding categorization of cognitive load has evolved over time. In the early stages when CLT was first proposed, cognitive load was categorized as either relating to schema construction (intrinsic cognitive load) or being unrelated to schema construction (extraneous cognitive load; Sweller et al., 1998). This is because early studies of cognitive load focused primarily on schema acquisition (Moreno & Park, 2010). Germane cognitive load was introduced in the 1990s when researchers found that partial cognitive load produced effects that were beneficial for learning (Sweller et al., 1998). It is capable of transferring knowledge to the cognitive load of long-term memory. In recent years, element interactivity has been expanded in research on cognitive load (Sweller, 2010), unifying the foundations of the three cognitive loads. The cognitive load discussed in this study uses a three-factor model: intrinsic cognitive load, extraneous cognitive load, and germane cognitive load (Sweller et al., 2019).

3.2.1 Intrinsic Cognitive Load

Intrinsic cognitive load (ICL) is determined by the complexity of the learning material and the learner's prior knowledge (Sweller, 1988; Leppink et al., 2013; Klepsch et al., 2017). The complexity of learning materials is related to the element interactivity. Materials with low element interactivity have each element that can be learned independently of the others, so even though there are many elements, they do not require much working memory. Materials with high element interactivity cannot be learned independently, and multiple elements must be considered simultaneously in the learning process. Therefore, the higher the element interactivity, the higher intrinsic load (Sweller, 2010). Additionally, the learner's prior knowledge is also a factor that influences the intrinsic load. If the learner already has a richer prior knowledge of the domain covered by the learning content, the new knowledge can be more quickly categorized into existing schemas, thus reducing the load on working memory (Sweller et al., 2019). However, more working memory is required to process more of the learning content when the learner's prior knowledge is insufficient, resulting in an increased intrinsic load (Leppink et al., 2014).

3.2.2 Extraneous Cognitive Load

Extraneous cognitive load (ECL) or ineffective cognitive load, which is related to how learning materials are presented and organized (Sweller et al., 2019). When learning materials are poorly designed, it can cause learners to unnecessarily process elements that are not relevant to learning (Sweller et al., 1998). For example, learners may be asked to unnecessarily search within materials for information to solve a problem or for an unclear reference in an explanation (Paas et al., 2003b). This can cause the learner to experience the split attention effect, resulting in an increase in extraneous load (Ayres & Sweller, 2005). Therefore, excessive extraneous load can interfere with learning and should be kept to a low level when designing instruction (Paas & Sweller, 2014).

3.2.3 Germane Cognitive Load

Germane cognitive load (GCL) is considered as a load necessary for learning (Schnotz & Kürschner, 2007). More cognitive resources are allocated to intrinsic load when a learner's extraneous cognitive load is low. This results in the processing of elements from working memory and their transfer to long-term memory (schema construction). The cognitive resources used in this process are called germane cognitive load (Sweller et al., 1998; Leppink et al., 2013). Thus, germane load is a cognitive resource needed to deal with intrinsic cognitive load (Sweller et al., 2019). However, it is also due to the very close relationship between GCL and ICL that the concept of germane load has been controversial since it was proposed. Some researchers have argued that germane load is not independent of the other two cognitive loads, but rather uses the same theoretical foundations as intrinsic load, making it indistinguishable from intrinsic load (Kalyuga, 2011). Other researchers have argued that GCL is an active load and that high GCL is a cognitive resource that learners invest in, whereas ICL is a load that is passively experienced (Moreno & Park, 2010; Klepsch & Seufert, 2021). In our study, we aim to investigate whether video tutorials are effective in transforming information into schemas, and therefore we will use the three-factor cognitive load categorization.

3.3 Effects of Video Tutorials

Video tutorials (VT) are a type of material that combines audio and visual information to provide step-by-step instruction that builds knowledge and skills (Noor et al., 2014; Ponzanelli et al., 2016). In recent years, the production, use, and effectiveness of video tutorials have gradually become a focus of research. From script development, video recording, audio recording, adding subtitles, to final editing and encapsulation (Oujezdský, 2014), the process of video tutorial production can involve a number of key steps. It is worth noting that although some researchers produced video tutorials based on cognitive load theory (Noor et al., 2013), CLT does not provide help on specific design problems (van der Meij & van der Meij, 2013). Therefore, many researchers have proposed some general design principles and considerations for the production of video tutorials (Martin & Martin, 2015; Nasir & Bargstädt, 2017; Fiorella & Mayer, 2018; Guy & McNally, 2022; van der Meij & Hopfner, 2022; Ring & Brahm, 2022). For example, van der Meij and van der Meij (2013) suggested eight guidelines for the design of video tutorials. The guidelines include many specific details about the design of video tutorials, such as recommending the use of highlighting to direct attention and the use of a conversational style to enhance perceptions of task relevance.

Some studies have examined the effects of video tutorials. A three-year study investigated the use of video tutorials in university programming courses (Wells et al., 2012). The video tutorials were introduced into the courses the first year without any modifications. The results showed that the satisfaction of the course increased, but 40% of the students still did not participate in the unit content. In the second year, researchers adjusted the video tutorials to better fit the assignments. This approach increased the number of assignments submitted and improved learning outcomes. In the third year, they repeated the methods from the second year and the results showed that 87% of the students used the tutorials to complete the assignments. However, the study also observed that face-to-face lectures were rated lower after the introduction of the video tutorials. Students preferred to use video tutorials for their learning.

Compared with other types of instructional materials, video tutorials have many advantages (Balslev et al., 2005; Zhang et al., 2006; Lloyd & Robertson, 2011; Gonzalves et al., 2018). In an empirical research, van der Meij and van der Meji (2014a) examined the difference between four types of instructional configuration: paper-based, paper-based preview and video procedure (Mixed A), video preview and paper-based procedure (Mixed B) and video tutorials. The 111 fifth and sixth-grade participants were randomly split into two groups. The result of post-test shows that the participants who used the video tutorials achieved the highest score of 76.8%, while the participants who used the paper-based materials only scored 55.6%. For the training tasks, the video tutorials also achieved the best success rate of 89.7%, while the paper-based materials resulted in a success rate of 65.4%. The results are consistent with the findings of Palmiter and Elkerton (1993), who found that use of video tutorials during training resulted in better final learning outcomes.

Many studies show that video tutorials can influence learners' cognitive load (Chen & Wu, 2015; Biard et al., 2018; Hughes et al., 2018). However, there are also contrary results.

Garrett (2018) conducted a study comparing the text, video, and segmented video in Excel learning. The 48 participants were randomly assigned to the three materials. The average completion time was 200 s for the video tutorial, 271 s for the segmented video, and 318 s for the text. Although the video tutorial was the most efficient format, scales showed no difference in the type of cognitive load utilized. In another study (Homer et al., 2008), 26 university students attended two different classrooms, one using slides with video tutorials, the other using slides without video tutorials. The cognitive load scales showed that students experienced higher cognitive load during instruction with video. However, the students achieved good learning outcomes in both classes. These results indicated that both materials are beneficial for learning, but the video tutorials required more mental effort.

In summary, although the effects of video tutorials have been widely researched, the empirical evidence on video tutorials compared with traditional instruction is still not clear. Therefore, our study aims to provide further evidence to clarify the impact of video instruction versus traditional instruction on cognitive load.

3.4 Methods of Cognitive Load Measurement

Effective measurement of cognitive load provides a basis for further study on video tutorials and is also a challenge in cognitive load research (de Jong, 2009; Ayres, 2017). It is now widely accepted that measures of cognitive load are categorized as subjective and objective (Brunken et al., 2003; Plass et al., 2010).

Subjective assessment based on learners' experiences and feelings during the learning process has been the main method of measuring cognitive load. Questions are usually related to the psychology of the learner and the difficulty of the task (Paas et al., 2003b; Schnotz & Kürschner, 2007). Measuring cognitive load from a subjective perspective was originally proposed by Paas (1992), and other researchers have developed a variety of methods to assess cognitive load. However, because the measurement dimensions are too simple or can only measure the total cognitive load, many researchers have tried to develop scales that measure one of the categories of cognitive load or different types of them (Ayres, 2006; DeLeeuw & Mayer, 2008; Cierniak et al., 2009; Leppink et al., 2013; Klepsch et al., 2017). The main advantage of subjective measurement methods is the convenience, requiring almost no instruments. The data obtained is also easy to analyze, and it does not interfere with the learners' learning tasks. However, it has some limitations. The results of the scales come from the subjective feelings of the learners, but sometimes the feelings can be different from the real mental load. Also, due to the controversy surrounding GCL, the subjective scale still needs to be further investigated.

In terms of objective measurement, dual-task paradigms are often used to measure participants' resource allocation status (Brunken et al., 2003). As technology advances, biofeedback techniques are also being used as an objective measurement and have been validated in many studies. For example, electroencephalography (EEG; Antonenko et al., 2010), functional magnetic resonance imaging (fMRI; Whelan, 2007), heart rate variability(HRV; Cranford et al., 2014; Minkley et al., 2018; Solhjoo et al., 2019), galvanic skin response (GSR; Conway et al., 2013; Larmuseau et al., 2019), and eye-tracking techniques (Recarte & Nunes, 2003; Van Gerven et al., 2004; Karch et al., 2019).

4 Method

4.1 Curriculum and Participants

We conducted a four-group controlled experiment in two vocational schools and one high school (two groups) in Germany. By having lessons from different types of schools and subjects, we were able to examine the effectiveness of the video tutorials in variety of teaching and learning environment. This enhanced the generalizability of the results.

The four groups were taught by four different teachers. Each teacher used a video tutorial in one of the lessons (video classroom: VC) and traditional teaching methods in another lesson (traditional classroom: TC). Teachers decided by themselves which lessons to use video tutorials and students were not aware of this in advance. In addition, the content of the lessons was strictly in line with the students' original study program. Only the development parts of the lessons were replaced by video tutorials in VC, all other parts followed the original lesson plan. Although the content may vary slightly, we have ensured that the materials are coherent and relevant as much as possible. For example, in Group A the learning topic was "Light and Color". In the video classroom, students first learned the basics knowledge of color vision, including how colors are perceived, color addition and color subtraction. In the traditional classroom, students learned the wavelength of light and how its reflection affects color. They also needed to understand the RGB and CMYK color spaces. All video tutorials were selected by the teacher based on course content and were taken from online resources. All students were able to control the pace of the videos while watching them.

Only the students who participated in both classes were selected as a sample to ensure that the observed cognitive load were not due to individual differences. In order to detect an effect size of Cohen's $d=0.50$ with 90% power ($\alpha=0.05$, two-tailed), G*Power 3.1 suggests we would need 44 participants in a paired samples t-test (Faul et al., 2009; Serdar et al., 2021). A total of 45 students (46.67% male, M_{age}= 19.42, SD_{age}= 2.54) fully participated in both class lessons. Group A was organized at a vocational school in Hamburg, Germany with the theme "Light and Color". A total of 15 students participated in both classes (20.00% male, M_{age} = 22.27; SD_{age} = 2.55). Group B took place in a vocational school in Elmshorn, Schleswig-Holstein, Germany, and the topic of the lessons was "Economy". A total of 5 students participated in both classes (100.00% male, M_{age} = 18.40; SD_{age} = 0.55). Group C and Group D took place in a high school in Ahrensburg, Schleswig-Holstein, Germany, with students from grade 12. The topic of the lesson in Group C was "Biology". A total of 13 students participated in both classes (61.54% male, M_{age} = 18.00; SD_{age} = 0.41). Group D's class topic was "Economics and Politics". A total of 12 of these students participated in both classes (41.67% male, M_{age} = 17.83; SD_{age} = 0.83). All participants were informed about the study in advance and signed an informed consent. Underage students had informed consent forms signed by their parents. All data were collected anonymously.

4.2 Subjective Cognitive Load Scale

In this study, we used the Cognitive Load Scale (CLS) to measure students' cognitive load. The CLS was chosen firstly because it is convenient and does not interfere with the students' learning process. Secondly, the CLS allows us to distinguish between three different types of cognitive load, which is important for exploring the differences between video and traditional classrooms. In addition, CLS has been demonstrated to be able to distinguish between complexity levels of the problems for perceived difficulty and mental effort (Ouwehand et al., 2021), and is also applicable to non-traditional instructions (Costley et al., 2020; Andersen & Makransky,2020).

We selected Leppink et al.'s (2013) Cognitive Load Scale. The CLS consists of ten items grouped into three dimensions, including ICL (three items, from 1 to 3), ECL (three items, from 4 to 6), and GCL (four items, from 7 to 10). The scale uses an 11-point Likert scale, where 0 means "not at all the case" and 10 means "completely the case". The reliability of the original version of the dimensions is Cronbach's α: ICL = 0.82, ECL = 0.75, and GCL = 0.82. In another meta-study that examined the reliability of various cognitive load scales, Leppink et al.'s scale also achieved good results (Cronbach's α: ICL = 0.845, ECL = 0.759, GCL = 0.909) and has been more widely used (Mutlu-Bayraktar et al., 2019; Krieglstein et al., 2022). Since the original version of the scale was used in a statistics course, the parts of the scale related to statistics were modified (see Fig. 1) and translated into German by a native speaker.

4.3 Procedure

Before the lesson began, four video cameras were placed and set up in the four corners of the classroom, and audio recorders were placed on the students' desks. Headphones were distributed to each student in the video classroom so that they would not be distracted by other students' videos. We also used the OBS software (version 27.2.3) to record students' screens in real time while they were engaged in video classroom. This allowed us to observe in retrospect how students interacted with the video tutorials, as well as their choices and strategies during the learning process. Participants were assigned a code as they entered the classroom. Participants were first given instructions by the researchers and then asked to turn on the audio recorder placed on their desks. At the same time, researchers turned on the video camera's recording mode. The researchers then left the classroom and the teacher began the lecture. At the end of the lecture, participants were asked to complete a cognitive load scale and simple demographic questions (age, gender).

All of the following questions refer to the activity (lecture, class, discussion session, skills training or study session) that just finished. Please respond to each of the questions on the following scale (0 meaning not at all the case and 10 meaning completely the case).
[1] The topic/topics covered in the activity was/were very complex.
[2] The activity covered contents that I perceived as very complex.
[3] The activity covered concepts and definitions that I perceived as very complex.
[4] The instructions and/or explanations during the activity were very unclear.
[5] The instructions and/or explanations were, in terms of learning, very ineffective.
[6] The instructions and/or explanations were full of un- clear language.
[7] The activity really enhanced my understanding of the topic(s) covered.
[8] The activity really enhanced my knowledge and understanding of the theme(s).
[9] The activity really enhanced my understanding of the context(s) covered.
[10] The activity really enhanced my understanding of concepts and definitions.

Fig. 1 Cognitive Load Scale

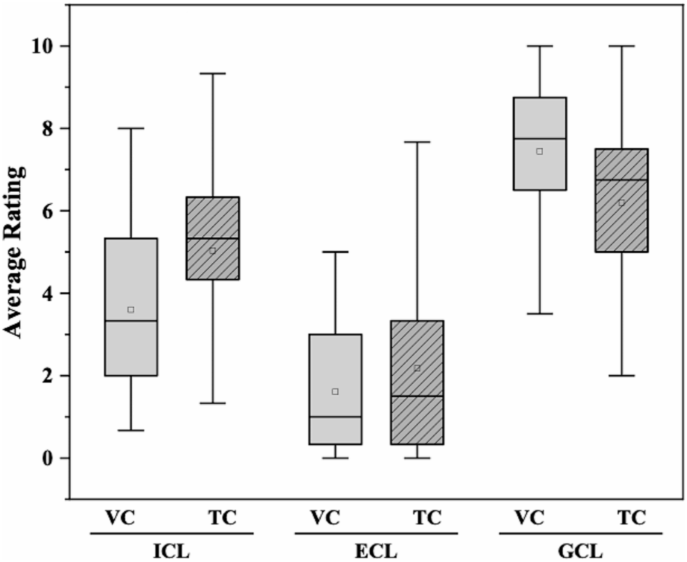


Fig. 2 Distributions of the three types of cognitive load for the video class versus the traditional class for the four groups combined. *Note N*=45. VC=Video classroom. TC=Traditional classroom

5 Results and Discussion

5.1 Differences in Cognitive Load for Four Groups

After checking the classroom of each group and selecting only the students who participated in both classes, 90 CLSs from 45 students were finally used for analysis. Initially, a normal distribution test using a quantile-quantile plot was performed on the difference between the two groups. The results showed that the data collected met the requirements for normal distribution. The ten items in the CLS were categorized according to cognitive load and dimensionality reduced for reliability analysis. The Cronbach's α: ICL = 0.892, ECL = 0.688, GCL = 0.898, which are within the acceptable range.

Figure 2 shows the distribution of the three cognitive loads for the four groups. The video classroom (M = 3.600, SD = 2.100) reports significantly lower scores on the ICL than the traditional classroom (M = 5.033, SD = 2.313), t = -4.507, $p < .001$, d = −0.672. Although the mean ECL is lower in the video classroom (M = 1.607, SD = 1.474) than in the traditional classroom (M = 2.174, SD = 2.245), there is no significant difference, t = -1.688, $p = .098$, d = −0.252. In terms of GCL, the video classroom (M = 7.444, SD = 1.681) reports significantly higher means than the traditional classroom (M = 6.200, SD = 1.786), t = 4.749, $p < .001$, d = 0.708).

These results answer research question 1, by establishing that there are differences in cognitive load between the video classroom and the traditional classroom. The results are consistent with previous studies (Chen & Wu, 2015; Biard et al., 2018; Hughes et al., 2018; Griffith & Faulconer, 2022), which indicated the video tutorials affect learners' cognitive load. Of the three types of cognitive load, students who completed video tutorials experienced significantly lower ICL and significantly higher GCL, while ECL showed almost no difference between the two classrooms.

Students in the video classroom reported lower ICL. According to CLT, instructional interventions cannot change ICL (Sweller, 1994). Changes in ICL are therefore influenced by the complexity of the interactive elements of the learning material and the learner's prior knowledge. In order to avoid individual differences, only participants who attended two classes were used in our results. Therefore, we can assume that there will not be much difference in students' prior knowledge of the topic when they participate in two classrooms. Thus, the element interactivity of the learning material can be considered here as the main factor influencing ICL. Researchers often express concerns that video tutorials may increase cognitive load due to the richness of the visual and auditory content (Paas et al., 2008). However, our results showed that students experience less ICL when using video tutorials. This is potentially because, although the video tutorial has sound and images, the elements involved in the knowledge are already combined and students do not have to spend much effort to re-process them. It has also been shown that videos with lower linguistic complexity produce lower ICL (Castro-Meneses et al., 2019). On the other hand, paper-based materials used in traditional classrooms require students to find what they need to learn and combine elements on their own, which can lead to more ICL.

In the ECL section, although the video classroom reported lower scores, there was no statistically significant difference between the two classrooms. However, it is noteworthy that ECL can be altered by instructional interventions (Van Merrienboer & Sweller, 2005). Therefore, additional attention needs to be paid to ECL when ICL is increased. Because the total cognitive load cannot exceed the capacity of working memory, when element interactivity is high, there is a need to focus on reducing ECL through instructional design (Paas et al., 2003b). One reason for the lack of difference in ECL between the two classrooms could be that the video tutorials gave the same clear instructions as the traditional classroom.

In terms of GCL, it is evident that learners experienced greater GCL in the video classroom. GCL is related to schema construction. High GCL is a sign that learners are actively transforming cognitive resources into schemas (Klepsch et al., 2017). This indicates that students in video classrooms invest more cognitive resources to process information. It also might show that video tutorials provide a more engaging learning experience than traditional classrooms. In addition, video tutorials offer flexibility. In the students' screen recordings, we found that students often watched a particular section repeatedly through the strategies by pausing and rewinding, which can contribute to deeper cognitive engagement. In contrast, the traditional classroom had a lower GCL. This means that although the traditional method was effective in transferring information, it was not as effective as video tutorials in promoting deep cognitive processes and meaningful learning.

It is important to note that although the video tutorial generally showed lower ICL and higher GCL, there may have been different results in different groups.

5.2 Differences in Cognitive Load for each Group

Because each classroom may produce different results due to differences in lesson design and teaching styles, we compare the differences in each of the four groups. Figure 3 shows the distributions of the three cognitive loads for Group A ($N=15$), Group B ($N=5$), Group

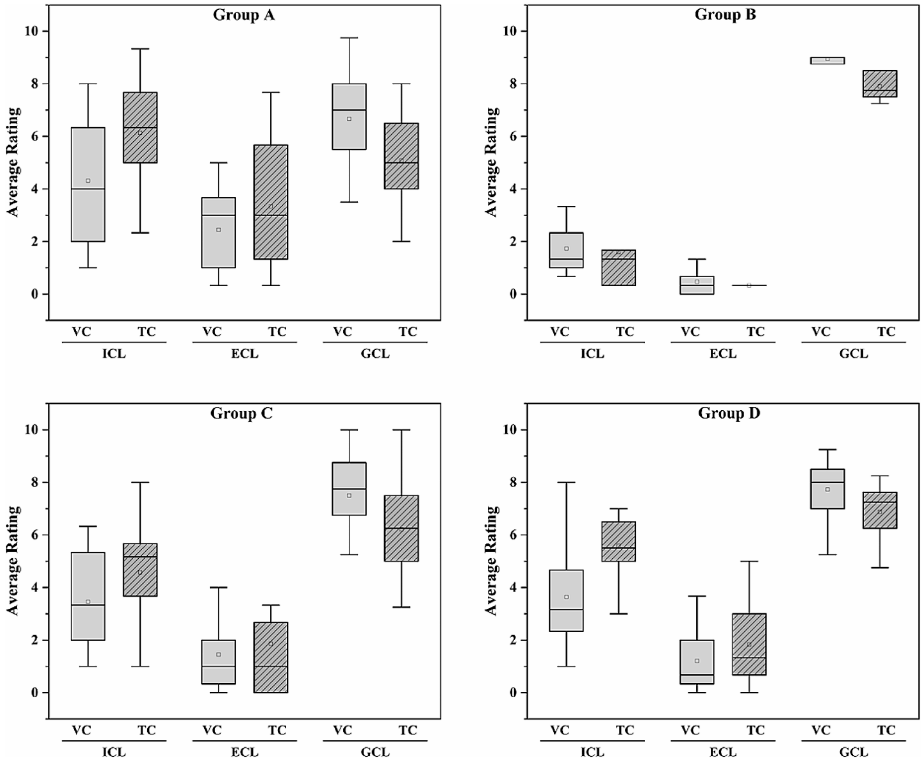


Fig. 3 Distributions of the three types of cognitive load in Group A, Group B, Group C and Group D for the video and traditional classrooms. *Note* Group A, $N=15$. Group B, $N=5$. Group C, $N=13$. Group D, $N=12$. VC = Video classroom. TC = Traditional classroom

Table 1 Paired t-test results of three types of cognitive load in Group A

	VC		TC		t (14)	*p*	Cohen's d
	M	SD	M	SD			
ICL	4.311	2.248	6.133	2.363	-2.815	0.014	−0.727
ECL	2.444	1.499	3.333	2.407	-1.186	0.255	−0.306
GCL	6.667	1.741	5.083	1.713	2.877	0.012	0.743

Note $N=15$. VC = Video classroom. TC = Traditional classroom

C ($N=13$) and Group D ($N=12$) for the video and traditional classroom. Tables 1, 2, 3 and 4 show the results of the paired t-test for the four groups.

From the results, it can be seen that Group A and Group C show the same results as the overall trend of significantly lower ICL and significantly higher GCL for the video classroom compared to the traditional classroom. Interestingly, Group B and Group D individually showed some differences to the overall cognitive load results. In Group B, the three cognitive loads barely differed between the two classrooms. One of the reasons for the lack of statistically significant differences in all items of Group B could be the small sample size ($N=5$). It is worth noting that students in Group D report lower ECL when learning with

Table 2 Paired t-test results of three types of cognitive load in Group B

	VC		TC		t (4)	p	Cohen's d
	M	SD	M	SD			
ICL	1.733	1.090	1.600	1.640	0.431	0.668	0.193
ECL	0.467	0.558	0.333	0.236	0.459	0.670	0.205
GCL	8.950	0.716	7.900	0.576	1.971	0.120	0.882

Note $N=5$. VC = Video classroom. TC = Traditional classroom

Table 3 Paired t-test results of three types of cognitive load in Group C

	VC		TC		t (12)	p	Cohen's d
	M	SD	M	SD			
ICL	3.462	1.989	4.577	2.028	-2.028	0.016	−0.777
ECL	1.449	1.478	1.859	2.468	−0.544	0.597	−0.151
GCL	7.500	1.920	6.211	1.928	2.646	0.021	0.734

Note $N=13$. VC = Video classroom. TC = Traditional classroom

Table 4 Paired t-test results of three types of cognitive load in Group D

	VC		TC		t (11)	p	Cohen's d
	M	SD	M	SD			
ICL	3.639	2.042	5.583	1.120	-2.737	0.019	−0.790
ECL	1.208	1.258	1.833	1.580	-2.245	0.046	−0.648
GCL	7.729	1.135	6.875	1.155	1.927	0.080	0.556

Note $N=12$. VC = Video classroom. TC = Traditional classroom

Table 5 Descriptive statistics and correlations for CL in Video Classroom

Variable	n	M	SD	1	2	3
1. ICL	45	3.60	2.10	—		
2. ECL	45	1.61	1.47	0.337^{*}	—	
3. GCL	45	7.44	1.68	-0.442^{**}	-0.332^{*}	—

$^*p<.05$. $^{**}p<.01$

Table 6 Descriptive statistics and correlations for CL in Traditional Classroom

Variable	n	M	SD	1	2	3
1. ICL	45	5.03	2.31	—		
2. ECL	45	2.17	2.24	0.498^{**}	—	
3. GCL	45	6.20	1.78	-0.451^{**}	-0.628^{*}	—

$^*p<.05$. $^{**}p<.01$

video tutorials. This means that the video tutorials in Group D could fit even more the students need. Further research it is necessary to analyze this specific difference.

5.3 Relationship between Three Types of Cognitive Load

Based on the results of the four groups, we can see that there are differences between video classroom and traditional classroom in terms of ICL and GCL. According to the additivity hypothesis of cognitive load, different types of cognitive load change dynamically when the total load does not exceed working memory capacity. The total cognitive load can be maintained by decreasing another type of cognitive load when one type of cognitive load increases. (Moreno & Park, 2010; Paas et al., 2003b). To explore the correlations between different types of cognitive load, we conducted a correlational analysis of the CLS results from each of the two classrooms (Research Question 2).

In Tables 5 and 6, we can see that the correlation between different types of cognitive load in both classrooms are similar. These results answer research question 2. ICL and ECL are in positive correlation, ICL and GCL are in negative correlation, and ECL and GCL also show negative correlation. According to the additivity hypothesis, when ICL levels are low, learners will have sufficient cognitive resources to deal with ECL. And when cognitive resources are progressively consumed by ECL, then fewer will be available for GCL (Park et al., 2015; Costley et al., 2020; Krieglstein et al., 2022). Our results show that ECL changes with ICL. In the video classroom, when the ICL is lower it leads to a higher GCL. And when ICL increased in the traditional classroom, leading to an increase in ECL, then GCL decreased. Changes in cognitive load between the two classrooms appear to be consistent with the additivity hypothesis. However, because of the limitations of subjective measures, and because we did not measure students' total load, we cannot determine whether students reached the limit of total load during the lesson. Therefore, our results can only suggest a potential feature of additivity across different types of cognitive load.

Meanwhile, to explore differences in correlations between different types of cognitive load in the two classrooms, we used Fisher's r-to-z transformation (Silver & Dunlap, 1987) to conduct a difference-in-difference analysis. As can be seen in Fig. 4, comparing ICL and ECL, the correlation for the video classroom is $z=0.397$, while the traditional classroom is $z=0.547$. The difference between the two classes is $z=-0.688$, which is not statistically significant, $p=.492$, $q=-0.150$. Comparing ICL and GCL, the correlation for the video classroom is $z=-0.475$ compared to $z=-0.486$ for the traditional classroom. The observed difference between the two classes is $z=0.052$, which is not statistically significant, $p=.959$, $q=0.011$. Comparing ECL and GCL, the correlation for the video classroom is $z=-0.345$, the correlation for the traditional classroom is $z=-0.738$. The difference between these correlations is $z=1.801$, $p=.072$, $q=0.393$. This suggests a trend towards statistical significance.

Based on the results of Fisher's r to z transformation, we can further answer research question 2. The ECL and GCL correlations in the two classrooms is approaching the border of significance ($z=1.801$, $p=.072$, $q=0.393$). ECL and GCL showed negative correlations in both the video classroom ($r=-.332$, $z=-0.345$) and the traditional classroom ($r=-.628$, $z=-0.738$), suggesting that GCL decreases as extraneous load increases. However, the decrease in germane load was greater in the traditional classroom compared to the video classroom. This means that in the traditional classroom, when there is an unfavorable ECL, there is a greater reduction in the cognitive resources available for schema construction

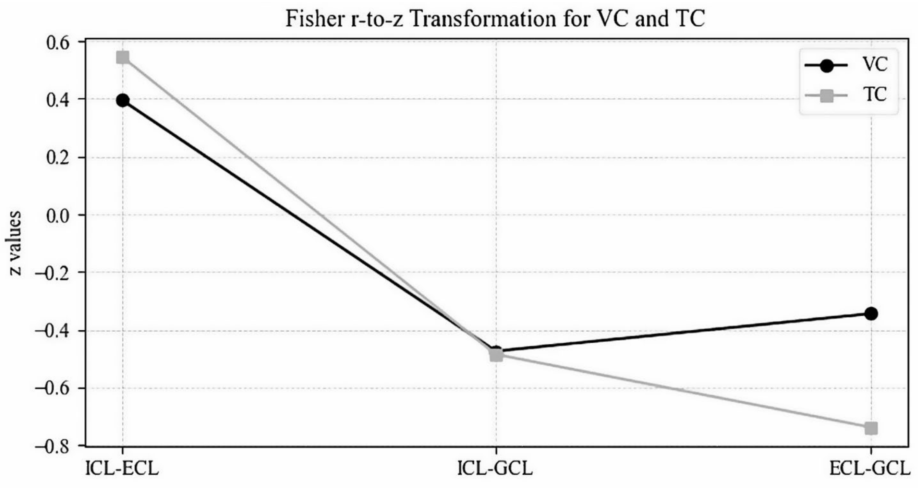


Fig. 4 Fisher r-to-z transformation for video classroom and traditional classroom

(GCL). The stronger negative correlations in the traditional classroom may be an indication that the overall cognitive load reaches its limits more often. As ECL increases, it takes away cognitive resources that could have been used for GCL.

6 Conclusions

Based on the results, we draw the following conclusions. There are differences in types of cognitive load experienced between video and traditional classrooms. In our study, students reported less ICL and increased GCL during the learning process using video tutorials compared to the traditional classroom. From the correlation analysis of the different types of cognitive load, it can be seen that learners adjust the use of other loads according to the use of different types of loads. This supports assertions that cognitive load is additive. As well, the correlations between ECL and GCL showed a trend of difference between the two classrooms. Learners were able to use the GCL to a greater extent when learning through video tutorials. Therefore, video tutorials appeared to be a favorable instructional material in this study. Not only did it involve less intrinsic load, but it also left more resources available for processing the germane load. These results demonstrate the potential of video tutorials to facilitate more effective and deeper learning.

7 Limitations and Future Work

At the same time, this study has several limitations. First, assessment of students' cognitive load used self-rating scales. This approach responds to students' subjective experiences, but not detailed information about specific cognitive processes. Therefore, in future studies, we will introduce objective measures such as heart rate variability to analyze cognitive load in more detail.

Second, the focus of this study was to measure the overall cognitive load of students in the classroom and not the specific details of the video tutorials. Although we recorded students' screens, we did not analyze them in detail. In future research, we could analyze the screen recordings in more depth. We could also add pre- and post-tests (Kulgemeyer et al., 2022) to examine whether students' viewing strategies created cognitive load and influenced learning outcomes. Eye tracking could also be used to more specifically analyze students' viewing strategies (Cook et al., 2017). For example, if students repeatedly use pause and rewind at some point in the video, is this related to their prior knowledge? Are students having a positive impact on learning outcomes by implementing viewing strategies?

Third, different instructional processes and learning steps may also affect changes in cognitive load. For example, ECL can be changed by instructional interventions (Van Merrienboer & Sweller, 2005), and teaching styles and instructional steps can also affect cognitive load. Therefore, students' classroom behaviors and teachers' organizational behaviors should be quantitatively coded in future research. Changes in cognitive load need to be analyzed from multiple perspectives including time on task, learning steps, social forms and motivation. We also suggest that similar studies in the future should include a follow-up survey of participants. For example, learning outcomes should be tested after one week to see how the knowledge is stored in long-term memory.

Fourth, although we controlled the variables in our study as much as possible, such as paired samples students, the same teacher in each group, strictly following the study program to ensure consistency and relevance of the teaching content, and the lesson design that only varies in the development part by using video tutorials or not, we still face some limitations. There may still be some slight differences in content, perhaps the video tutorials covered more complex or simpler content in one group than in the traditional classroom. And each teacher's approach may vary from classroom to classroom, even within the same instructional framework. Such differences may be due to adjustments based on student responses in the classroom, or due to different strategies in different instructional settings.

Finally, although the results observed changes in different cognitive loads and the potential for additivity, a more comprehensive study is required to draw firmer conclusions about the additivity hypothesis. And as learners progress in the classroom, it is possible that the intrinsic load decreases over time while other loads change accordingly. This fluidity implies that cognitive load is not static during the learning process, and therefore it is necessary to add dynamic studies of cognitive load in the future (Leppink et al., 2013; Paas et al., 2016). Furthermore, the present study was conducted under the assumption of the existence of germane cognitive load, whereas there are still many questions about the concept of germane cognitive load (de Jong, 2009). Germane cognitive load is difficult to distinguish from intrinsic cognitive load, and more research is required to explore and clarify the framework of cognitive load theory.

Author Contributions All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Enqi Fan and Jens Siemon. The first draft of the manuscript was written by Enqi Fan. Matt Bower, Jens Siemon revised it critically for important intellectual content. All authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.
Open Access funding enabled and organized by Projekt DEAL.

References

Andersen, M. S., & Makransky, G. (2020). The validation and further development of a multidimensional cognitive load scale for virtual environments. *Journal of Computer Assisted Learning*, *37*(1), 183–196. https://doi.org/10.1111/jcal.12478

Anmarkrud, Ø., Andresen, A., & Bråten, I. (2019). Cognitive load and working memory in Multimedia Learning: Conceptual and measurement issues. *Educational Psychologist*, *54*(2), 61–83. https://doi.org/10.1080/00461520.2018.1554484

Antonenko, P., Paas, F., Grabner, R., & van Gog, T. (2010). Using Electroencephalography to measure cognitive load. *Educational Psychology Review*, *22*(4), 425–438. https://doi.org/10.1007/s10648-010-9130-y

Ayres, P. (2006). Using subjective measures to detect variations of intrinsic cognitive load within problems. *Learning and Instruction*, *16*(5), 389–400. https://doi.org/10.1016/j.learninstruc.2006.09.001

Ayres, P. (2017). Subjective measures of cognitive load. *Cognitive Load Measurement and Application*, 9–28. https://doi.org/10.4324/9781315296258-2

Ayres, P., & Sweller, J. (2005). The Split-attention Principle in Multimedia Learning. *The Cambridge Handbook of Multimedia Learning*, 135–146. https://doi.org/10.1017/cbo9780511816819.009

Balslev, T., de Grave, W. S., Muijtjens, A. M., & Scherpbier, A. J. (2005). Comparison of text and video cases in a postgraduate problem-based learning format. *Medical Education*, *39*(11), 1086–1092. https://doi.org/10.1111/j.1365-2929.2005.02314.x

Bartlett, F. C. (1932). *Remembering: A study in experimental and social psycology*. University.

Biard, N., Cojean, S., & Jamet, E. (2018). Effects of segmentation and pacing on procedural learning by video. *Computers in Human Behavior*, *89*, 411–417. https://doi.org/10.1016/j.chb.2017.12.002

Bower, G. H. (2014). Cognitive psychology: An introduction. In *In Handbook of Learning and Cognitive Processes (Volume 1)* (pp. 25–80). essay, Psychology Press.

BROADBENT, D. E., & IMMEDIATE MEMORY AND THE SHIFTING OF ATTENTION. (1958). *Perception and Communication*, 210–243. https://doi.org/10.1016/b978-1-4832-0079-8.50011-6

Brown, J. (1958). Some tests of the Decay Theory of Immediate Memory. *Quarterly Journal of Experimental Psychology*, *10*(1), 12–21. https://doi.org/10.1080/17470215808416249

Brunken, R., Plass, J. L., & Leutner, D. (2003). Direct measurement of cognitive load in Multimedia Learning. *Educational Psychologist*, *38*(1), 53–61. https://doi.org/10.1207/s15326985ep3801_7

Castro-Meneses, L. J., Kruger, J. L., & Doherty, S. (2019). Validating theta power as an objective measure of cognitive load in educational video. *Educational Technology Research and Development*, *68*(1), 181–202. https://doi.org/10.1007/s11423-019-09681-4

Chen, C. M., & Wu, C. H. (2015). Effects of different video lecture types on sustained attention, emotion, cognitive load, and learning performance. *Computers & Education*, *80*, 108–121. https://doi.org/10.1016/j.compedu.2014.08.015

Chi, M., Glaser, R., & Rees, E. (1982). Expertise in problem solving. In R. Sternberg (Ed.), *Advances in the psychology of human intelligence* (pp. 7–75). Erlbaum.

Cierniak, G., Scheiter, K., & Gerjets, P. (2009). Explaining the split-attention effect: Is the reduction of extraneous cognitive load accompanied by an increase in germane cognitive load? *Computers in Human Behavior*, *25*(2), 315–324. https://doi.org/10.1016/j.chb.2008.12.020

Conway, D., Dick, I., Li, Z., Wang, Y., & Chen, F. (2013). The effect of stress on cognitive load measurement. *Human-Computer Interaction – INTERACT*, *2013*, 659–666. https://doi.org/10.1007/978-3-642-40498-6_58

Cook, A. E., Wei, W., & Preziosi, M. A. (2017). The use of ocular-motor measures in a convergent approach to studying cognitive load. *Cognitive Load Measurement and Application*, 112–128. https://doi.org/10.4324/9781315296258-8

Costley, J., Fanguy, M., Lange, C., & Baldwin, M. (2020). The effects of video lecture viewing strategies on cognitive load. *Journal of Computing in Higher Education*, *33*(1), 19–38. https://doi.org/10.1007/s12528-020-09254-y

Cranford, K. N., Tiettmeyer, J. M., Chuprinko, B. C., Jordan, S., & Grove, N. P. (2014). Measuring load on Working Memory: The Use of Heart Rate as a Means of Measuring Chemistry Students' cognitive load. *Journal of Chemical Education*, *91*(5), 641–647. https://doi.org/10.1021/ed400576n

de Jong, T. (2009). Cognitive load theory, educational research, and instructional design: Some food for thought. *Instructional Science*, *38*(2), 105–134. https://doi.org/10.1007/s11251-009-9110-0

de Koning, B. B., Tabbers, H. K., Rikers, R. M., & Paas, F. (2007). Attention cueing as a means to enhance learning from an animation. *Applied Cognitive Psychology*, *21*(6), 731–746. https://doi.org/10.1002/acp.1346

DeLeeuw, K. E., & Mayer, R. E. (2008). A comparison of three measures of cognitive load: Evidence for separable measures of intrinsic, extraneous, and germane load. *Journal of Educational Psychology*, *100*(1), 223–234. https://doi.org/10.1037/0022-0663.100.1.223

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. https://doi.org/10.3758/brm.41.4.1149

Fiorella, L., & Mayer, R. E. (2018). What works and doesn't work with instructional video. *Computers in Human Behavior*, *89*, 465–470. https://doi.org/10.1016/j.chb.2018.07.015

Ganier, F., & de Vries, P. (2016). Are instructions in video format always better than photographs when learning manual techniques? The case of learning how to do sutures. *Learning and Instruction*, *44*, 87–96. https://doi.org/10.1016/j.learninstruc.2016.03.004

Garrett, N. (2018). Twenty-fourth Americas Conference on Information Systems.

Gonzalves, A., Verhaeghe, C., Bouet, P. E., Gillard, P., Descamps, P., & Legendre, G. (2018). Effect of the use of a video tutorial in addition to simulation in learning the maneuvers for shoulder dystocia. *Journal of Gynecology Obstetrics and Human Reproduction*, *47*(4), 151–155. https://doi.org/10.1016/j.jogoh.2018.01.004

Griffith, J., & Faulconer, E. (2022). Show Me! Do Videos Make a Difference in An Asynchronous Online Course? *Online Journal of Distance Learning Administration, 25*(2). Retrieved from https://commons.erau.edu/publication/1820

Guy, J., & McNally, M. B. (2022). Ten key factors for making Educational and instructional videos. *Scholarly and Research Communication*, *13*(2). https://doi.org/10.22230/src.2022v13n2a423

Hamas, A., Maryono, D., & Efendi, A. (2019). The effectiveness of the use of blended learning model based on video tutorial reviewed from students' understanding in the Basic Graphic Learning in Banyudono State Vocational School. *Journal of Informatics and Vocational Education*, *2*(2). https://doi.org/10.20961/joive.v2i2.37972

Hogarth, K., & Luke, B. (2010). Transitioning from dependent to more independent learners: Using video tutorials as a means of helping students help themselves. In *Proceedings of the 2010 Accounting and Finance Association of Australia and New Zealand (AFAANZ) Conference* (pp. 1-2rizz1). Accounting and Finance Association of Australia and New Zealand (AFAANZ).

Homer, B. D., Plass, J. L., & Blake, L. (2008). The effects of video on cognitive load and social presence in multimedia-learning. *Computers in Human Behavior*, *24*(3), 786–797. https://doi.org/10.1016/j.chb.2007.02.009

Hughes, C., Costley, J., & Lange, C. (2018). The effects of self-regulated learning and cognitive load on beginning to watch and completing video lectures at a cyber-university. *Interactive Technology and Smart Education*, *15*(3), 220–237. https://doi.org/10.1108/itse-03-2018-0018

Käfer, V., Kulesz, D., & Wagner, S. (2016). *What Is the Best Way for Developers to Learn New Software Tools? A Small Empirical Comparison between a Text and a Video Tutorial*. https://doi.org/10.7287/peerj.preprints.2413v2

Kalyuga, S. (2011). Cognitive load theory: How many types of load does it really need? *Educational Psychology Review*, *23*(1), 1–19. https://doi.org/10.1007/s10648-010-9150-7

Kalyuga, S., & Sweller, J. (2014). The redundancy Principle in Multimedia Learning. *The Cambridge Handbook of Multimedia Learning*, 247–262. https://doi.org/10.1017/cbo9781139547369.013

Karch, J. M., García Valles, J. C., & Sevian, H. (2019). Looking into the Black Box: Using gaze and Pupillometric Data to Probe how cognitive load changes with Mental tasks. *Journal of Chemical Education*, *96*(5), 830–840. https://doi.org/10.1021/acs.jchemed.9b00014

Kirschner, P. A. (2002). Cognitive load theory: Implications of cognitive load theory on the design of learning. *Learning and Instruction*, *12*(1), 1–10. https://doi.org/10.1016/s0959-4752(01)00014-7

Klepsch, M., & Seufert, T. (2021). Making an Effort Versus Experiencing Load. *Frontiers in Education*, *6*. https://doi.org/10.3389/feduc.2021.645284

Klepsch, M., Schmitz, F., & Seufert, T. (2017). Development and validation of two instruments measuring intrinsic, extraneous, and germane cognitive load. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.01997

Krieglstein, F., Beege, M., Rey, G. D., Ginns, P., Krell, M., & Schneider, S. (2022). A systematic Meta-analysis of the reliability and validity of subjective cognitive load questionnaires in experimental Multimedia Learning Research. *Educational Psychology Review*, *34*(4), 2485–2541. https://doi.org/10.1007/s10648-022-09683-4

Kulgemeyer, C., Hörnlein, M., & Sterzing, F. (2022). Exploring the effects of physics explainer videos and written explanations on declarative knowledge and the illusion of understanding. *International Journal of Science Education*, *44*(11), 1855–1875. https://doi.org/10.1080/09500693.2022.2100507

Larkin, J. H., McDermott, J., Simon, D. P., & Simon, H. A. (1980). Models of competence in solving physics Problems*. *Cognitive Science*, *4*(4), 317–345. https://doi.org/10.1207/s15516709cog0404_1

Larmuseau, C., Vanneste, P., Cornelis, J., Desmet, P., & Depaepe, F. (2019). Combining physiological data and subjective measurements to investigate cognitive load during complex learning. *Frontline Learning Research*, *7*(2), 57–74. https://doi.org/10.14786/flr.v7i2.403

Leppink, J., & Van den Heuvel, A. (2015). The evolution of cognitive load theory and its application to medical education. *Perspectives on Medical Education*, *4*(3), 119–127. https://doi.org/10.1007/s40037-015-0192-x

Leppink, J., Paas, F., Van der Vleuten, C. P., Van Gog, T., & Van Merriënboer, J. J. (2013). Development of an instrument for measuring different types of cognitive load. *Behavior Research Methods*, *45*(4), 1058–1072. https://doi.org/10.3758/s13428-013-0334-1

Leppink, J., Paas, F., van Gog, T., van der Vleuten, C. P. M., & van Merriënboer, J. J. G. (2014). Effects of pairs of problems and examples on task performance and different types of cognitive load. *Learning and Instruction*, *30*, 32–42. https://doi.org/10.1016/j.learninstruc.2013.12.001

Lloyd, S. A., & Robertson, C. L. (2011). Screencast tutorials Enhance Student Learning of statistics. *Teaching of Psychology*, *39*(1), 67–71. https://doi.org/10.1177/0098628311430640

Luke, B., & Hogarth, K. (2011). Developing and enhancing independent learning skills. *Accounting Research Journal*, *24*(3), 290–310. https://doi.org/10.1108/10309611111187019

Martin, P. A. (2016). Tutorial video use by senior undergraduate electrical engineering students. *Australasian Journal of Engineering Education*, *21*(1), 39–47. https://doi.org/10.1080/22054952.2016.1259027

Martin, N. A., & Martin, R. (2015). Would you watch it? Creating effective and engaging video tutorials. *Journal of Library & Information Services in Distance Learning*, *9*(1–2), 40–56. https://doi.org/10.1080/1533290x.2014.946345

Mayer, R. E. (2005). Cognitive Theory of Multimedia Learning. *The Cambridge Handbook of Multimedia Learning*, 31–48. https://doi.org/10.1017/cbo9780511816819.004

Mayer, R. E., Fiorella, L., & Stull, A. (2020). Five ways to increase the effectiveness of instructional video. *Educational Technology Research and Development*, *68*(3), 837–852. https://doi.org/10.1007/s11423-020-09749-6

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81–97. https://doi.org/10.1037/h0043158

Minkley, N., Kärner, T., Jojart, A., Nobbe, L., & Krell, M. (2018). Students' mental load, stress, and performance when working with symbolic or symbolic–textual molecular representations. *Journal of Research in Science Teaching*, *55*(8), 1162–1187. https://doi.org/10.1002/tea.21446

Morain, M., & Swarts, J. (2012). YouTutorial: A Framework for assessing Instructional Online Video. *Technical Communication Quarterly*, *21*(1), 6–24. https://doi.org/10.1080/10572252.2012.626690

Moreno, R., & Park, B. (2010). Cognitive load theory: Historical development and relation to other theories. *Cognitive Load Theory*, 9–28. https://doi.org/10.1017/cbo9780511844744.003

Mutlu-Bayraktar, D., Cosgun, V., & Altan, T. (2019). Cognitive load in multimedia learning environments: A systematic review. *Computers & Education*, *141*, 103618. https://doi.org/10.1016/j.compedu.2019.103618

Nasir, A. R., & Bargstädt, H. J. (2017). An Approach to develop video tutorials for construction tasks. *Procedia Engineering*, *196*, 1088–1097. https://doi.org/10.1016/j.proeng.2017.08.066

Noor, N. M., Hamizan, N. I., & Rahim, R. A. (2013). The framework for learning using video based on cognitive load theory among visual learners. *2013 IEEE 5th Conference on Engineering Education (ICEED)*. https://doi.org/10.1109/iceed.2013.6908295

Noor, N. M., Aini, M., & Hamizan, N. I. (2014). Video based learning embedded with cognitive load theory: Visual, auditory, and kinaesthetic learners' Perspectives. *2014 International Conference on Teaching and Learning in Computing and Engineering*. https://doi.org/10.1109/latice.2014.19

Oujezdský, A. (2014). Creation of Educational Video tutorials and their use in Education. *International Journal of Information and Communication Technologies in Education*, *3*(1), 28–39. https://doi.org/10.1515/ijicte-2014-0003

Ouwehand, K., van Kroef, A., Wong, J., & Paas, F. (2021). Measuring cognitive load: Are there more valid alternatives to likert rating scales? *Frontiers in Education, 6*. https://doi.org/10.3389/feduc.2021.702616

Paas, F. G. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology*, *84*(4), 429–434. https://doi.org/10.1037/0022-0663.84.4.429

Paas, F., & Sweller, J. (2014). Implications of cognitive load theory for Multimedia Learning. *The Cambridge Handbook of Multimedia Learning*, 27–42. https://doi.org/10.1017/cbo9781139547369.004

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and Instructional Design: Recent developments. *Educational Psychologist*, *38*(1), 1–4. https://doi.org/10.1207/s15326985ep3801_1

Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. (2003b). Cognitive Load Measurement as a Means to Advance cognitive load theory. *Educational Psychologist*, *38*(1), 63–71. https://doi.org/10.1207/s15326985ep3801_8

Paas, F. G. W. C., Ayres, P., & Pachman, M. (2008). Assessment of cognitive load in multimedia learning. *Recent Innovations in Educational Technology that Facilitate Student Learning Information Age Publishing Inc Charlotte*, *NC*, 11–35.

Paas, F., Renkl, A., & Sweller, J. (Eds.). (2016). *Cognitive load theory: A special issue of educational psychologist*. Routledge.

Palmiter, S., & Elkerton, J. (1993). Animated demonstrations for learning procedural computer-based tasks. *Human-Computer Interaction*, *8*(3), 193–216. https://doi.org/10.1207/s15327051hci0803_1

Park, B., Korbach, A., & Brünken, R. (2015). Do learner characteristics moderate the seductive-details-effect? A cognitive-load-study using eye-tracking. *Journal of Educational Technology & Society*, *18*(4), 24–36.

Peterson, L., & Peterson, M. J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology*, *58*(3), 193–198. https://doi.org/10.1037/h0049234

Plass, J. L., Kalyuga, S., & Leutner, D. (2010). Individual Differences and Cognitive Load Theory. *Cognitive Load Theory*, 65–88. https://doi.org/10.1017/cbo9780511844744.006

Ponzanelli, L., Bavota, G., Mocci, A., Di Penta, M., Oliveto, R., Hasan, M., Russo, B., Haiduc, S., & Lanza, M. (2016). Too long; didn't watch! Extracting relevant fragments from software development video tutorials. *Proceedings of the 38th International Conference on Software Engineering*. https://doi.org/10.1145/2884781.2884824

Recarte, M. A., & Nunes, L. M. (2003). Mental workload while driving: Effects on visual search, discrimination, and decision making. *Journal of Experimental Psychology: Applied*, *9*(2), 119–137. https://doi.org/10.1037/1076-898x.9.2.119

Ring, M., & Brahm, T. (2022). A rating framework for the quality of video explanations. *Technology Knowledge and Learning*. https://doi.org/10.1007/s10758-022-09635-5

Rizza, M., Widho Wati, C., & Dardiri, A. (2019). Transforming Digital Learning in Vocational High School 21st Century. *Proceedings of the 2nd International Conference on Vocational Education and Training (ICOVET 2018)*. https://doi.org/10.2991/icovet-18.2019.65

Rozi, F., Hamid, A., K., & Panjaitan, K. (2020). Development of Tutorial Video Media Based on Project Based Learning in Class XI State Vocational School 1 Pakam Lubuk. *Journal of Physics: Conference Series*, *1485*(1), 012056. https://doi.org/10.1088/1742-6596/1485/1/012056

Rumelhart, D. E. (1980). On evaluating Story Grammars*. *Cognitive Science*, *4*(3), 313–316. https://doi.org/10.1207/s15516709cog0403_5

Schnotz, W., & Kürschner, C. (2007). A reconsideration of cognitive load theory. *Educational Psychology Review*, *19*(4), 469–508. https://doi.org/10.1007/s10648-007-9053-4

Serdar, C. C., Cihan, M., Yücel, D., & Serdar, M. A. (2021). Sample size, power and effect size revisited: Simplified and practical approaches in pre-clinical, clinical and laboratory studies. *Biochemia Medica*, *31*(1), 27–53. https://doi.org/10.11613/bm.2021.010502

Silver, N. C., & Dunlap, W. P. (1987). Averaging correlation coefficients: Should Fisher's z transformation be used? *Journal of Applied Psychology*, *72*(1), 146–148. https://doi.org/10.1037/0021-9010.72.1.146

Solhjoo, S., Haigney, M. C., McBee, E., van Merrienboer, J. J., Schuwirth, L., Artino, A. R., Battista, A., Ratcliffe, T. A., Lee, H. D., & Durning, S. J. (2019). Heart Rate and Heart Rate Variability correlate with clinical reasoning performance and self-reported measures of cognitive load. *Scientific Reports*, *9*(1). https://doi.org/10.1038/s41598-019-50280-3

Sweller, J. (1988). Cognitive load during problem solving: Effects on Learning. *Cognitive Science*, *12*(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*, *4*(4), 295–312. https://doi.org/10.1016/0959-4752(94)90003-5

Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational Psychology Review*, *22*(2), 123–138. https://doi.org/10.1007/s10648-010-9128-5

Sweller, J., van Merrienboer, J. J., & Paas, F. G. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, *10*(3), 251–296. https://doi.org/10.1023/a:1022193728205

Sweller, J., Ayres, P., & Kalyuga, S. (2011). Measuring cognitive load. *Cognitive Load Theory*, 71–85. https://doi.org/10.1007/978-1-4419-8126-4_6

Sweller, J., van Merriënboer, J. J., & Paas, F. (2019). Cognitive Architecture and Instructional Design: 20 years later. *Educational Psychology Review*, *31*(2), 261–292. https://doi.org/10.1007/s10648-019-09465-5

van der Meij, H., & Hopfner, C. (2022). Eleven guidelines for the design of instructional videos for Software Training. *Technical Communication*, *69*(3), 5–23. https://doi.org/10.55177/tc786532

van der Meij, H., & van der Meij, J. (2013). Eight guidelines for the design of instructional videos for Software Training. *Technical Communication*, *60*(3), 205–228.

van der Meij, H., & van der Meij, J. (2014a). A comparison of paper-based and video tutorials for software learning. *Computers & Education*, *78*, 150–159. https://doi.org/10.1016/j.compedu.2014.06.003

van der Meij, J., & van der Meij, H. (2014b). A test of the design of a video tutorial for software training. *Journal of Computer Assisted Learning*, *31*(2), 116–132. https://doi.org/10.1111/jcal.12082

van der Meij, H., & van der Meij, J. (2016). The effects of reviews in video tutorials. *Journal of Computer Assisted Learning*, *32*(4), 332–344. https://doi.org/10.1111/jcal.12136

Van Gerven, P. W., Paas, F., Van Merrienboer, J. J., & Schmidt, H. G. (2004). Memory load and the cognitive pupillary response in aging. *Psychophysiology*, *41*(2), 167–174. https://doi.org/10.1111/j.1469-8986.2003.00148.x

van Merrienboer, J. J., & Sweller, J. (2005). Cognitive load theory and Complex Learning: Recent developments and future directions. *Educational Psychology Review*, *17*(2), 147–177. https://doi.org/10.1007/s10648-005-3951-0

Wahyudi, S., Joyoatmojo, S., & Sawiji, H. (2017). Learning model of attention, relevance, confidence, satisfaction (ARCS) supported by Video Tutorial to improve the students' learning motivation in Vocational High School. *Proceedings of the International Conference on Teacher Training and Education 2017 (ICTTE 2017)*. https://doi.org/10.2991/ictte-17.2017.72

Wells, J., Barry, R. M., & Spence, A. (2012). Using video tutorials as a Carrot-and-Stick Approach to Learning. *IEEE Transactions on Education*, *55*(4), 453–458. https://doi.org/10.1109/te.2012.2187451

Whelan, R. R. (2007). Neuroimaging of cognitive load in instructional multimedia. *Educational Research Review*, *2*(1), 1–12. https://doi.org/10.1016/j.edurev.2006.11.001

Zhang, D., Zhou, L., Briggs, R. O., & Nunamaker, J. F. (2006). Instructional video in e-learning: Assessing the impact of interactive video on learning effectiveness. *Information & Management*, *43*(1), 15–27. https://doi.org/10.1016/j.im.2005.01.004

Zinn, B., Tenberg, R., & Pittich, D. (2021). Erklärvideos – Im Naturwissenschaftlich-Technischen Unterricht Eine Alternative zu Texten? [explanatory videos as an alternative to text-based learning using the problem-solving methodology SPALTEN as an example?]. *Journal of Technical Education*, *9*(2), 168–187. https://doi.org/10.48513/joted.v9i2.217

Computers & Education

Comparing cognitive load during video versus traditional classroom instruction based on heart-rate variability measures

Enqi Fan[a*], Matt Bower[b], Jens Siemon[a]

[a]*Department of Vocational and Business Education, University of Hamburg, Sedanstraße 19, 20146, Hamburg, Germany*
[b]*School of Education, Macquarie University, Building 29WW, Balaclava Rd North Ryde, 2109, NSW Sydney, Australia*

Abstract

This pilot study used heart rate variability (HRV) as an indicator of cognitive load to examine student engagement in the learning process. We investigated the dynamics of students' (N=45, paired sample) cognitive load in classes with and without video tutorials and compared differences in cognitive load between development phases of lessons where students are acquiring new knowledge. The results of the study show that the average cognitive load of students is higher when using video tutorials than in classrooms without them. From the students' behavior, we can see that when using video tutorials, students frequently adjust their viewing strategies or take notes. In classrooms without videos, students are more easily distracted. This means that students mobilized more cognitive resources for effective learning while using video tutorials. In general, our results suggest that the use of video tutorials in the development phase of classroom can increase student effectiveness in learning new knowledge. This study provides new insights into the application of video tutorials as a form of computer-assisted instruction, highlighting the potential benefits of using dynamic cognitive load monitoring in real classroom environments.

Keywords: video tutorial, lesson phases, heart rate variability, dynamics cognitive loady

1. Introduction

Video tutorials (VT), as a form of computer-assisted instruction, have become increasingly popular in recent years (Sharma, 2017). Video tutorials present learning content more vividly through a combination of sound and images compared to traditional instructional materials (Hong et al., 2016). In addition, learners can control their learning progress through different viewing strategies such as pausing and rewinding (Costley et al., 2020; Kuhlmann et al., 2024; Luke & Hogarth, 2011; Murphy et al., 2022), which allows them to actively engage in the learning process. However, when students watch videos, are they truly learning? Do video tutorials really have an advantage over traditional teaching materials? While several studies have indicated that the use of video tutorials in teaching has a positive impact on student engagement (Ding et al., 2024; Lloyd & Robertson, 2011; Van der Meij & Van der Meij, 2014; Wahyudi et al., 2017;

*Corresponding author. *E-mail address:* enqi.fan@uni-hamburg.de
Submitted: July 2024

Wells et al., 2012), this is not always the case. Factors such as the length, type, style of the video, and the viewing strategies can significantly influence engagement (Anders et al., 2024; Chen & Thomas, 2020; Deng, 2024; Parker et al., 2024). In recent studies, student engagement has often been measured by analyzing video logs, including pause frequency, viewing counts, and skipped segments (Kuhlmann et al., 2024; Liao & Wu, 2023; Parker et al., 2024). Although these data can indicate students' engagement behaviors, they may not accurately reflect the cognitive engagement of the learners. Students may exhibit high frequencies of viewing and varied playback strategies, but this does not necessarily equate to a deep understanding of the material (Akçapınar & Bayazıt, 2024; Tseng, 2021). This gap between behavioral engagement and cognitive understanding presents a significant challenge in educational research, emphasizing the necessity of measuring engagement from a cognitive perspective.

Cognitive load theory provides a good framework for analyzing whether students are truly engaged in learning (Sweller, 1988; Sweller, 2023). When learners are less engaged, they use relatively few cognitive resources in the learning process. In contrast, as learners become more engaged, they use more cognitive resources to undertake deeper processing (Miller, 2015). Therefore, by measuring students' cognitive load, teachers can know how much effort is put into knowledge processing at the cognitive level and what content is more likely to contribute to effective learning. There have been many studies investigating the effects of video tutorials on cognitive load (Biard et al. 2018; Chen & Wu, 2015; Costley & Lange, 2017; Hughes et al., 2018; Wyeld, 2016). However, most studies have focused on measuring overall cognitive load or a specific type of cognitive load (Costley & Lange, 2017). For example, Fan et al. (2024) investigated the effect of video tutorials on different types of cognitive load in traditional classrooms using participant self-report responses to the Cognitive Load Scale. It is also worth noting that cognitive load is variable during the course of learning (Leppink et al., 2013; Paas et al., 2016). However, previous studies have focused mainly on static and holistic measurements of cognitive load across an entire learning episode and have not fully examined the dynamic changes in it during the learning process, for instance using biometric approaches. Students might experience a high or low cognitive load when instruction begins. This cognitive load changes as time passes. This means that there may be different cognitive loads at different phases of instruction. A class often has different lesson phases due to the teacher's instructional design. For example, organization phases that are not related to the teaching content, introduction phases that are related to the topic, development phases for learning new knowledge, practice phases for applying the knowledge, or debriefing phases where the teacher and the students discuss the results of the learning. Therefore, examining cognitive load at different lesson phases can help teachers better understand how their students are learning. Comparing video tutorials and traditional instruction in specific phases could give information about which is more effective.

Therefore, in this study, we introduce heart rate variability (HRV, Mullikin et al., 2024) to indicate students' cognitive load levels in different lesson phases. Unlike subjective measurements such as questionnaires or scales, biofeedback technology can more objectively reflect students' dynamic cognitive changes. It is worth noting that this study only explores HRV in depth during the "development phase" of the classroom. This is because the development phase focuses on the process of students learning new knowledge. The direct impact of video tutorials on students during this phase may be the most significant. By studying the changes in cognitive load during this key stage, it is possible to gain insight into how video tutorials can promote students' cognitive engagement in understanding and mastering new knowledge. The results of the study will provide evidence of the effectiveness of video tutorials in terms of cognitive engagement, which will help teachers gain a deeper understanding of students' use of cognitive resources. At the same time, it will fill the research gap in comparing dynamic cognitive load in a real classroom environment.

1.1. Cognitive load and engagement

In educational research, cognitive load is often used as an indicator to examine learners' use of cognitive resources. This is because cognitive load theory indicates that learners need to use cognitive resources to transform information into schemas if they want to store knowledge in long-term memory (Kirschner, 2002; Paas et al., 2003; Rumelhart, 1980). There are three different types of cognitive load: intrinsic, extraneous, and germane. The intrinsic load depends on the inherent difficulty of the material and the learner's pre-existing knowledge. Extraneous load is influenced by how the instructional content is structured and delivered. Germane load, on the other hand, pertains to the development of schemas (Kalyuga, 2011; Sweller et al., 1998).

There is an association between learners' cognitive load and engagement (Berka et al., 2007; Bueno-Vesga et al., 2021). Some studies have demonstrated through subjective scales that higher cognitive load is often accompanied by higher engagement (Lan et al. 2019; Zheng et al., 2023). In Wu et al. (2022)'s study, it was found through EEG (electroencephalography) measurements that learners' engagement increased along with the level of mental load. The richness of video tutorials can also have an impact on cognitive load when used in the classroom (Paas et al., 2008). Many researchers are concerned that video tutorials may increase cognitive load due to the rich visual and auditory content (Mayer, 2005; Paas et al., 2008). However, video tutorials also increase germane cognitive load and facilitate schema construction (Costley & Lange, 2017; Costley et al., 2021; Fan et al., 2024). This increase in cognitive load is actually beneficial to learning, which indicates that the learner is transforming knowledge into a schema (Kalyuga, 2011).

1.2. Measuring cognitive load with heart rate variability

Effective measurement of cognitive load has been an ongoing focus of cognitive load research (Ayres, 2017; de Jong, 2009). Besides intrinsic, extraneous and germane cognitive load, cognitive load can also be divided into different levels: instantaneous load, peak load, average load, accumulated load and overall load (Xie & Salvendy, 2000). Instantaneous load is a dynamic manifestation of cognitive load, which reflects the changing cognitive demand of the learner at each moment during the task. Peak load indicates the maximum value of instantaneous load during the task. Accumulated load measures the total cognitive resource consumption during the task. The average load, on the other hand, represents the average level of cognitive load over the duration of the task. In addition, the overall load describes the learner's subjective sense of mental effort over the entire task and is usually measured by a scale (Paas et al., 2003). Compared with instantaneous and peak loads, average and accumulated loads are considered key indicators for evaluating the effectiveness of teaching interventions. Because they can reveal the relationship between task duration and cognitive resource consumption (Antonenko et al., 2010).

Subjective measures and objective measures are two main ways to measure cognitive load (Brunken et al., 2003; Plass et al., 2010). From one dimension to multiple dimensions, subjective scales have been widely used to measure different types of cognitive load (Ayres, 2006; Cierniak et al., 2009; DeLeeuw & Mayer, 2008; Dönmez et al.,2022; Klepsch et al., 2017; Paas, 1992).

Although the subjective measures are easier to obtain and to analyze, they still have some limitations (Chen et al., 2011). For example, the differences between different scales (Klepsch et al., 2017) and the way in which the scale is scored (Ouwehand et al., 2021) can have impacts on the measurement of cognitive load. In addition, subjective scales cannot measure the dynamic changes in cognitive load (Antonenko et al., 2010). In contrast, biofeedback methods are increasingly being applied to measure cognitive load. For example, EEG (electroencephalography) can measure cognitive load (Antonenko et al., 2010; Baceviciute et al.,2022; Makransky et al.,2019; Örün & Akbulut, 2019), and fMRI (functional Magnetic Resonance Imaging) technology enables all three types of cognitive load to be directly measured based on areas of brain activity (Whelan, 2007). However, the burdensome constraints of the experimental environment and the high price make fMRI and EEG unsuitable for classroom research. Instead, researchers prefer smaller, portable, and non-invasive devices that can measure biomarkers, such as HRV (heart rate variability; Cranford et al., 2014; Minkley et al., 2018; Solhjoo et al., 2019), GSR (galvanic skin responses; Conway et al., 2013; Larmuseau et al., 2019), and eye-movements (Karch et al., 2019; Souchet et al., 2022; Šola et al., 2024). Among these, wearable HRV devices have been highly favored by researchers due to their portability, non-invasiveness, and low cost.

Our heart responds to changes in the environment and to the needs of the body, and these responses appear as slight variations in the continuous cycle of heartbeats. This is due to the automatic nervous system (ANS), which controls the heart's activity (Ziemssen & Siepmann, 2019). Physiological arousal occurs when people interact with their environment, which influences changes in the ANS (Levenson, 2003). The ANS is closely related to changes in a person's subjective emotional experience, which is divided into the sympathetic nervous system (SNS) and the parasympathetic nervous system (PNS; Egger et al., 2019). The SNS will dominate during periods of high activity and stress, resulting in an increase in heart rate (HR) and a decrease in heart rate variability (HRV). The PNS will dominate during quiet and relaxed states, resulting in a decrease in HR and an increase in HRV. (Pham et al.,2021). Therefore, HRV can be used as an objective indicator to assess the state of the ANS (Bernardi et al., 2000; Hjortskov et al., 2004; Mullikin et al., 2024; Rolim et al., 2013). It has been shown that there is a correlation between HRV and cognitive load. When cognitive load is increased, it leads to higher blood pressure and more frequent breathing, which can lead to a decrease in HRV (Arutyunova et al., 2024; Grassmann et al., 2015; Solhjoo et al., 2019; Song & Lehrer, 2003). Therefore, there have been many studies using HRV as an indicator of cognitive load (Haapalainen et al., 2010; Minkley et al., 2018; Mizuno et al., 2011; Mukherjee et al., 2011).

In this study, we will objectively indicate cognitive load using HRV. Although it is not possible to distinguish between intrinsic, extraneous, and germane cognitive load, we can acquire an indication of average load at different lesson phases and show the instantaneous load changes during the development phase.

1.3. Present study

In this study, we will measure students' cognitive load in both traditional classroom (TC) and video classroom (VC). We focus on the development phase of the lesson where students are acquiring new knowledge because this phase most commonly introduces video tutorials to replace the original traditional teaching methods. In order to directly measure the changes in students' cognitive load, we used heart rate variability (HRV) as a biofeedback indicator. We hypothesize that video tutorials will increase students' cognitive load and thus promote effective learning, and therefore propose the research question:

RQ: Do students who use video tutorial compared to traditional instruction have a difference in cognitive load during the development phase of learning?

2. Method

2.1. Course Content and Participants

A controlled experiment was conducted with four groups from two vocational schools and one high school in Germany. By including lessons from diverse school types and subjects, we evaluated the performance of VTs in different instructional contexts, which strengthened the applicability of our findings.

Four teachers, each assigned to a group, instructed using both video tutorials (VC: video classroom) and traditional methods (TC: traditional classroom). The teachers independently selected which lessons would incorporate the video content, without informing the students in advance. To respect the original lesson plan, we let the teachers decide on their own which lesson to use the video tutorial in first. Groups A, B, and C all used the video tutorial in the first recorded lesson. Group D's teacher used the video tutorial in the second recorded lesson. The lesson content strictly followed the students' standard curriculum, with only the development parts substituted by video tutorials in the video classroom. Everything else remained consistent with the original curriculum. Although the materials varied slightly, efforts were made to ensure coherence and relevance across the lessons. For example, Group A covered the topic of "Unemployment". In the first lesson, the teacher presented the unemployment data and briefly identified the types of unemployment. Students then began to work independently through the video tutorials to further understand the types, causes, and effects of unemployment. In the second lesson, the teacher reviewed the content of the previous lesson. Students then engaged in individual work by reading the text to understand the issue of unemployment financial assistance. Teachers selected all video tutorials according to the course material and sourced them from online platforms. In group A, the teacher provided two different video tutorials. Video A is 3:47 long and introduces the main reasons and types of unemployment in Germany. Video B is 2:07 long and introduces the German unemployment benefit system. Both videos are narrated animations with highlighted key points. In group B, the teacher provided one video, which is 12:53 minutes long. In this video, a Bauen&Wohnen worker and a police officer walk into an apartment and explain in detail the function of a mechanical security system. In group C, the teacher provided two videos. Video A is 2:03 minutes long and explains the process of meiosis in the human body through narrated animation. Video B is 3:26 long and explains the structure of chromosomes and the inheritance of certain characteristics in a narrated animation. In group D, the teacher provided one video that is 9:12 long. A presenter in the video explains the concept of "Federal Europe" and people's opinions about it, along with the animation. Figure 1 shows screenshots of the videos used in the four groups. As students watched the videos, they could modify the playback speed.

We only included students who participated in both classes as the sample for analysis, to prevent individual differences from influencing the measurement of cognitive load. A priori power analysis using G*Power 3.1 indicated that a paired sample of at least 44 participants was required to detect a large effect size with a power of 0.90, and alpha of 0.05 (Faul et al., 2009). 45

students (21 males, Mage= 19.87, SDage= 3.66) took part in both the VC and TC. All students signed informed consents, with parents signing on behalf of underage students. All data collection was anonymous. Basic information about the four classroom groups and participants is shown in Table 1.

Table 1. Group Information and Participant Demographics

Group	School Type	Topic	$N_{students}$	N_{male}	M_{age}	SD_{age}
A	Vocational school	Unemployment	10	1	23	3.94
B	Vocational school	Security	10	7	21.6	4.9
C	High school	Biology	13	8	18	0.41
D	High school	Economics and Politics	12	5	17.83	0.83

Fig. 1. *Screenshots of the video tutorials used in four groups (Group A top left, Group B top right, Group C bottom left, Group D bottom right)*

2.2. HRV capture

We used the Polar H10 heart rate band (Polar Electro Oy, Kempele, Finland) as the HRV collection device. It is a small and portable heart rate band that has been shown to correctly measure RR intervals (Gilgen-Ammann et al., 2019; Speer et al., 2020; Moya-Ramon et al., 2022). The Polar H10 heart rate band was sampled at a frequency of 1000 Hz, and data were transferred via Bluetooth connection to the iMotions software (9.3, iMotions, Copenhagen, Denmark) in four high performance laptops. All data were stored locally.

2.3. Procedure

We positioned cameras in the four corners of the classroom to observe the class from multiple perspectives. Each participant also wore a voice recorder. In the VC, we gave each participant a pair of headphones, which prevented them from being bothered by others' playback. Participants were assigned a code and a heart rate band as they entered the classroom. The researchers instructed the participants to first put on the heart rate band and then to activate the voice recorder. Meanwhile, a researcher activated the filming mode. After that, all researchers exited the classroom and the teaching began. Participants needed to fill out a brief demographic questionnaire (age, gender) before leaving the classroom.

2.4. *Video and Audio Analysis*

We used the video analysis software Mangold interact (Mangold International GmbH, Arnstorf, Germany) to code the classroom videos. This is a qualitative and quantitative video coding software. It can play multiple videos and audio simultaneously. The researchers coded the classroom videos according to the coding manual "Classroom Observation Coding Manual System - Lesson Phases, Social Forms, and Time-on-Task" (Author, in press; details see Appendix A). The coding manual categorized and defined the following lesson steps.

Code 1. Learning Organization: The teacher organizes activities which are not related to the content of the lesson topic.

Code 2. Introduction: The teacher starts to talk about the course topic but has not yet communicated the new content of the lesson.

Code 3. Development: Let students learn and familiarize with brand new content.

Code 4. Practice: Students apply what they have already learned.

Code 5. Debriefing: Communication between the teacher and students regarding the results of the lesson.

Code a. Learning Organization (intervention): The teacher organizes activities which are not related to the content of the lesson.

Code b. Introduction (intervention): The teacher starts to talk about the course topic but has not yet communicated the new content of the lesson.

Code c. Development (intervention): Let students learn and familiarize with brand new content through interventions.

Code d. Practice (intervention): Students apply previous learning through interventions.

Code e. Debriefing (intervention): Checking or comparing student solutions through intervention.

Code 9. Mixed form: This categorization is unlikely to happen in the classroom

Code 0. not assignable: This is also unlikely to happen.

The coding manual is intended to cover the lesson steps that may occur, as well as a few special cases (e.g., Code 9, Code 0). The Kappa value of the coding manual is 0.71-1.00 (McHugh, 2012). The coding process was completed by multiple trained researchers who thoroughly coded the video and audio data according to the coding manual.

2.5. HRV Analysis

We exported raw csv data from the iMotions software. Before further calculations of HRV, we cropped the heart rate data outside the classroom time by checking the video timestamps. Outliers in RR intervals were removed using the 3sigma rule. We used HRVanalysis (version: 1.0.4) to analyze the HRV data. This is a Python module for HRV analysis, originally developed by the R&D team at OCTO Technology as part of the Aura Healthcare project. It started development in July 2018 and is distributed under the GNU General Public License Version 3 (GPLv3) license (Champseix et al., 2021).

There are several methods for analyzing HRV, such as time domain analysis, frequency domain analysis, and nonlinear dynamics analysis. Since we need to calculate HRV for different phases of the lesson, the duration of these phases varies from a few seconds to several minutes. Therefore, we chose RMSSD from time domain analysis as an indicator of cognitive load (Stein et al., 1994). RMSSD is the root mean square of successive differences between normal heartbeats, and it can be calculated even in the case of short heart rate recordings (Thong et al., 2003; Pham et al., 2021). RMSSD is an indicator of parasympathetic nervous system. Therefore, when the RMMSD value decreases it means higher cognitive load. However, because RMSSD values can be affected by individual differences in participants' health, we first calculated RMSSD mean values for each participant in each classroom. We then calculated the percentage difference in RMSSD from the mean for each student across the lesson steps. We also calculated the percentage change in RMSSD per minute for each student. Percentage changes allowed us to observe dynamic changes in students' cognitive load.

3. Results and Discussion

This section first shows the RMSSD differences in overall development phases in the four groups of classrooms. It then shows the different lesson steps durations and RMSSD percentage changes for each classroom. We have also plotted the instantaneous load during the development phase. In the tables and figures, the development phases in VC is highlighted in blue, and in TC is highlighted in red. A positive RMSSD percentage means that the students' cognitive load is below the classroom average. A negative RMSSD percentage means that the students' cognitive load is above the classroom average.

3.1. Difference in cognitive load at development phases

We examined the generalization of cognitive load by combining data from all students. Before the analysis, we used the Quantile-Quantile (Q-Q) plot to check the normality of the paired data differences. This showed that the data satisfied a normal distribution. We first performed a paired t-test on the overall data of 45 paired samples, and then performed separate paired t-tests on four groups. To control the risk of Type I Error in multiple tests, we used the Bonferroni correction. At the same time, to avoid Type II Error, we also used the Holm-Bonferroni correction. Table 2 shows the paired t-test result of the development part RMSSD percentage changes in VC and TC.

Table 2. *Results of paired t-test analysis of Development part in four groups*

	VC		TC		t (44)	p	Bonferroni	Holm	Cohen's d
	M	SD	M	SD					
RMSSD (%)	-4.213	8.808	2.512	7.183	-4.205	<.001	<.001	<.001	-.627

As can be seen from the Table 2, in VC the cognitive load of students in the development phase is generally higher than average. In TC, on the other hand, students' cognitive load in the development phase is generally lower than average. Significance of the overall result remains after using Bonferroni and Holm-Bonferroni corrections. This means that more cognitive load is generated when students use video tutorials to learn new content. This means that students may be more engaged when using video tutorials. It is worth mentioning that, even though VC typically had a greater CL, results could vary across the groups.

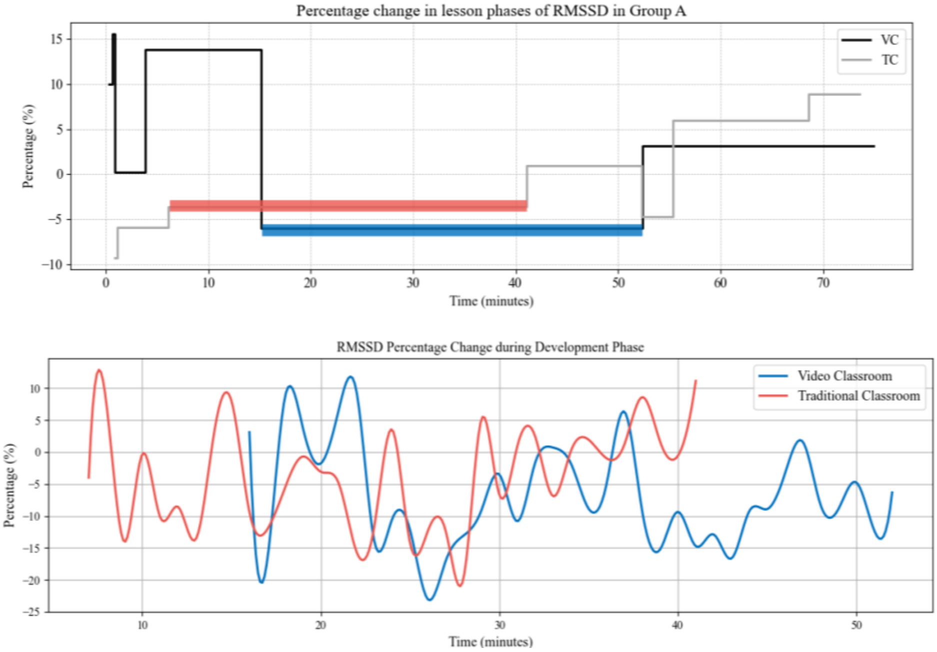
3.2. Differences in cognitive load for lesson steps in Group A

Table 3 shows the duration (min) and percentage change of different lesson steps of Group A in VC and TC. A visual comparison of the percentage change in RMSSD of lesson steps for Group A in VC and TC is shown in Figure 2. Table 4 shows the paired t-test result of the development part RMSSD percentage changes in VC and TC.

Table 3. *Lesson steps time and percentage change value of RMSSD in Group A*

VC				TC			
Code		Time (min)	RMSSD (%)	Code		Time (min)	RMSSD (%)
2	Introduction	0.38	9.942	1	Organization	0.26	-9.323
1	Organization	0.25	15.581	2	Introduction	5.00	-5.885
2	Introduction	2.92	0.231	**3**	**Development**	34.91	**-3.621**
3	Development	11.30	13.788	5	Debriefing	11.24	0.917
c	**Development (VT)**	37.27	**-5.979**	1	Organization	3.07	-4.705
5	Debriefing	22.54	3.150	4	Practice	13.21	5.998
				5	Debriefing	4.94	8.914

Fig. 2. *Percentage change of RMSSD in Group A*



Note. Development phases are highlighted.

Table 4. *Results of paired t-test analysis of Development part in Group A*

	VC		TC		t (9)	p	Bonferroni	Holm	Cohen's d
	M	SD	M	SD					
RMSSD (%)	-5.979	6.702	-3.621	5.584	-.712	.495	1.000	.494	-.225

In Group A, the teacher conducted the development phase in VC for 37.27 minutes and the mean RMSSD percentage for students was -5.979%. In TC, the development phase was conducted for 34.91 minutes and the mean RMSSD percentage for students was -3.621%. Although the duration of the development phase was similar in both lessons and the RMSSD percentages were both below the mean, there was no statistical difference. However, as shown in Figure 2, the mean percentage of RMSSD was lower for students in VC. This means that students generated slightly more cognitive load while learning new content using the video tutorials. The RMSSD change curve over the development phase shows that RMSSD fluctuates several times at the beginning of the VC development phase. As can be seen in the classroom recordings, at the beginning the students operate the tablet, connect headphones and have some brief exchanges with their peers. Afterwards, the students watch the video independently and complete the learning task. We can notice that the students are very focused when watching the video, often using viewing strategies such as pausing or rewinding, and rarely communicate with their peers. In TC, the RMSSD level of the students is very similar to VC. In the classroom recordings, we can see that students complete their learning tasks by reading PDF materials. However, during this time, some students often discuss with their peers, which may cause interference to other students who are learning independently.

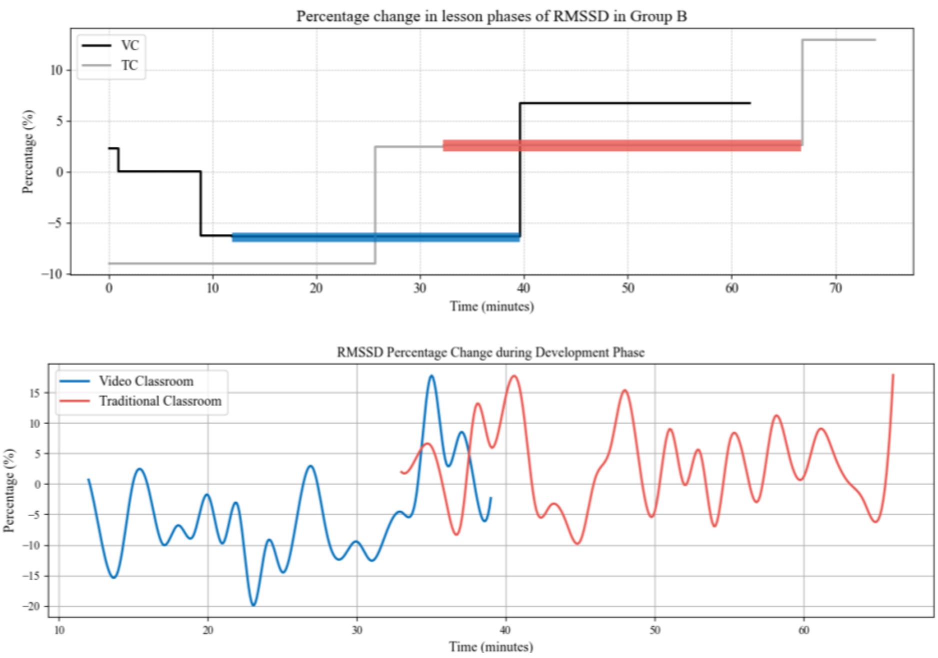
3.3. Differences in cognitive load for lesson steps in Group B

Table 5 shows the duration (min) and percentage change of different lesson steps of Group B in VC and TC. A visual comparison of the percentage change in RMSSD of lesson steps for Group B in VC and TC is shown in Figure 3. Table 6 shows the paired t-test result of the development part RMSSD percentage changes in VC and TC.

Table 5. *Lesson steps time and percentage change value of RMSSD in Group B*

VC				TC			
Code		Time (min)	RMSSD (%)	Code		Time (min)	RMSSD (%)
1	Organization	0.89	2.349	2	Introduction	25.60	-9.006
2	Introduction	7.95	0.060	1	Organization	6.58	2.476
1	Organization	2.98	-6.261	**3**	**Development**	34.56	**2.669**
c	**Development (VT)**	27.81	**-6.328**	5	Debriefing	7.05	13.030
5	Debriefing	22.11	6.802				

Fig. 3. *Percentage change of RMSSD in Group B*



Note. Development phases are highlighted.

Table 6. *Results of paired t-test analysis of Development part in Group B*

	VC		TC		t (9)	p	Bonferroni	Holm	Cohen's d
	M	SD	M	SD					
RMSSD (%)	-6.328	10.715	2.669	8.223	-1.934	.085	.425	.170	-.612

In group B, the teacher conducted the development phase in VC for 27.81 minutes and the mean RMSSD percentage of the students was -6.328%. The development phase was conducted in TC for 34.56 minutes and the mean RMSSD percentage for students was 2.669%. Teachers used more time to cover new content in TC, but students' RMSSD was above the classroom average. This means that in the development phase of TC, students did not invest many cognitive resources. In VC, on the other hand, students' cognitive load is higher than in TC, even though there was no statistically significant difference. In the RMSSD percentage curve, we can see that the RMSSD of the students in the beginning of the video tutorial is consistently below the average of the class. In the last few minutes of the development phase, the RMSSD of the students begins to rise. In the classroom recordings, we can see that the students are indeed very attentive at the beginning of the development phase. The students watch the video on their tablets and occasionally pause to take notes. During this process, the students rarely interact with their peers. Considering that Group B used the documentary video, which contains dialogues between the two characters over time, the students needed to pay more attention to catch the main points. In the later stage of the development phase, the students completed the video viewing and learning tasks, and their RMSSD showed a noticeable increase. In the traditional classroom, the teacher arranged for the students to read the material together in groups, and it can be seen that the students' RMSSD fluctuated around the average level.

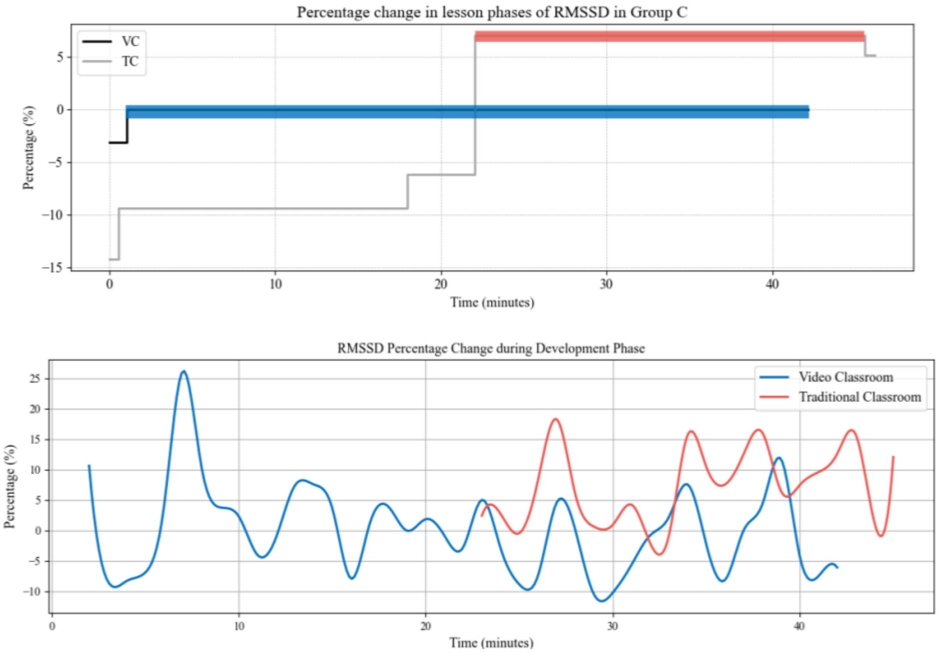
3.4. Differences in cognitive load for lesson steps in Group C

Table 7 shows the duration (min) and percentage change of different lesson steps of Group C in VC and TC. A visual comparison of the percentage change in RMSSD of lesson steps for Group C in VC and TC is shown in Figure 4. Table 8 shows the paired t-test result of the development part RMSSD percentage changes in VC and TC.

Table 7. *Lesson steps time and percentage change value of RMSSD in Group C*

VC				TC			
Code		Time (min)	RMSSD (%)	Code		Time (min)	RMSSD (%)
2	Introduction	1.08	-3.133	1	Organization	0.55	-14.236
c	**Development (VT)**	41.06	**0.038**	5	Debriefing	17.45	-9.388
				1	Organization	4.04	-6.194
				3	**Development**	23.52	**7.059**
				1	Organization	0.61	5.130

Fig. 4. *Percentage change of RMSSD in Group C*



Note. Development phases are highlighted.

Table 8. *Results of paired t-test analysis of Development part in Group C*

	VC		TC		t (12)	p	Bonferroni	Holm	Cohen's d
	M	SD	M	SD					
RMSSD (%)	.038	.619	7.059	7.497	-3.460	.005	.023	.018	-.960

In group C, teachers conducted the development phase in VC for 41.06 minutes and the mean value of RMSSD percentage of students was 0.038%. In TC, the development phase was conducted for 23.52 minutes and the mean value of RMSSD percentage of students was 7.059%. In VC, the teacher spent only one minute to set up the learning task and then the students started learning independently. The development phase took almost the whole class. This means that the teacher gave the students the freedom to learn independently using the video tutorials. We found that students' RMSSD was higher than average when using the video tutorials. This suggests that students' cognitive load is slightly lower in the development phase. However, when we compare it with TC, we found that the percentage of RMSSD is significantly higher in TC's development phase. This means that students in TC invest much less cognitive resources in learning. The curve shows that at the beginning of the VT development phase, students' RMSSD showed a significant peak. Through the classroom recordings, we found that during this phase, most students were operating their laptops to log in to the learning system. When students started watching the video, RMSSD began to drop and fluctuate within a small range. During this phase, students watched the video alone and completed the learning tasks. As can be seen from the screen recordings, after watching the video, students often dragged the progress bar to a certain point and watched it again. In TC, the students' RMSSD was almost constantly above the average for the development phase. The students completed the learning tasks by reading the material, but they communicated with each other more frequently than in VC.

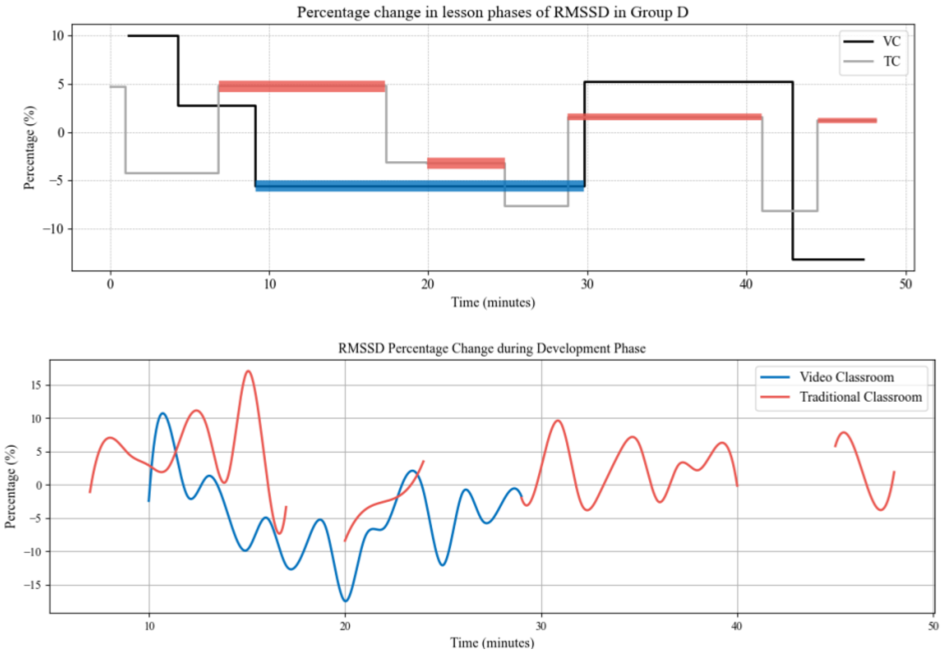
3.5. Differences in cognitive load for lesson steps in Group D

Table 9 shows the duration (min) and percentage change of different lesson steps of Group D in VC and TC. A visual comparison of the percentage change in RMSSD of lesson steps for Group D in VC and TC is shown in Figure 5. Table 10 shows the paired t-test result of the development part RMSSD percentage changes in VC and TC.

Table 9. *Lesson steps time and percentage change value of RMSSD in Group D*

VC				TC			
Code		Time (min)	RMSSD (%)	Code		Time (min)	RMSSD (%)
2	Introduction	3.09	10.056	1	Organization	0.95	4.771
6	Introduction (VT)	4.89	2.743	2	Introduction	5.83	-4.205
c	**Development (VT)**	20.68	**-5.586**	**3**	**Development**	10.57	**4.822**
5	Debriefing	13.06	5.252	5	Debriefing	2.53	-3.110
4	Practice	4.43	-13.173	**3**	**Development**	4.88	**-3.143**
				5	Debriefing	3.98	-7.629
				3	**Development**	12.23	**1.559**
				5	Debriefing	3.45	-8.128
				3	**Development**	3.69	**1.254**

Fig.5. Percentage *change of RMSSD in Group D*



Note. Development phases are highlighted.

Table 10. *Results of paired t-test analysis of Development part in* Group *D*

	VC		TC		t (11)	p	Bonferroni	Holm	Cohen's d
	M	SD	M	SD					
RMSSD (%)	-5.586	12.204	2.566	2.358	-2.683	.021	.106	.063	-.775

In group D, teachers conducted the development phase in VC for 20.68 minutes and the mean value of RMSSD percentage for students was -5.586%. A total of 31.37 minutes of development phases were conducted in TC and the mean value of RMSSD percentage for students was 2.566%. In VC, the cognitive load was higher in the development phase. From the curve graph, we can notice that after the students entered the development phase, RMSSD showed a short-term increase, and then it was in a downward trend. Classroom recordings show that students watch videos independently and take notes during this phase. In TC, we can see that the teacher conducted four individual development phases out of all the lesson phases. The teacher used both explanation and independent reading of the material during the four development phases. The RMSSD percentages in these four development phases were different, but generally higher than in VC. This implies that students still invested fewer cognitive resources in the development phases in TC.

3.6. Discussion

Based on the above results, we can answer the research question that there is a difference in cognitive load in the development phase of video classroom and traditional classroom. The results are consistent with previous studies (Biard et al., 2018; Chen & Wu, 2015; Hughes et al., 2018), which indicated the video tutorials affect learners' cognitive load. As can be seen from the students' behavior in VC and the changes in RMSSD, students' high-frequency viewing strategies are indeed accompanied by more cognitive engagement (Kuhlmann et al., 2024; Liao & Wu, 2023; Parker et al., 2024). Classroom recordings show that students tend to be more attentive when they use video tutorials for learning, while there are more distractions in traditional classrooms. This indicates that video tutorials can improve student engagement (Ding et al., 2024; Lloyd & Robertson, 2011; Van der Meij & Van der Meij, 2014; Wahyudi et al., 2017; Wells et al., 2012).

It is worth noting that the uniqueness of our study is that it shows the changes in cognitive load during different phases in a real classroom environment through HRV. We can clearly observe in the figures from Group A to Group D that students' cognitive load varies with teachers' lesson phases. These rarely exist in previous studies. However, it should be noted that although HRV can indicate both average and instantaneous load, it cannot distinguish between the types of cognitive load. This means that we need to consider whether high cognitive load is affected by extraneous load. Although some researchers are concerned that too high cognitive load can create a burden for learning (Mayer, 2005; Paas et al., 2008), it is not necessarily negative. It has been found in several studies that high cognitive load is often followed by high engagement (Lan et al., 2019; Zheng et al., 2023). In addition, a study found no difference between the extrinsic load of video and traditional classrooms using a subjective scale, but that students had more germane load when using videos (Fan et al., 2024). Germane load is a factor that promotes schema construction (Costley & Lange, 2017; Kirschner, 2002; Paas et al., 2003). Through behavioral observation, we can see that when using videos, students frequently take notes or adjust their viewing strategies, which is a sign of active engagement in learning. In traditional classrooms, students engage in peer interactions more frequently. These results support the definition of germane load in cognitive load theory, which states that learners allocate working memory to learning-related activities (Krieglstein et al., 2023; Sweller, 2010). Therefore, we can assume that students used their cognitive resources more reasonably during the development phase with video tutorials. This high cognitive load responds to the fact that students actively mobilized the germane cognitive load when using video tutorials and transformed their knowledge into schemas.

Another point of interest is that although the mean RMSSD percentage in the VC is lower in all four groups, we can note that there is no statistically significant difference between groups A and B. However, there is either a significant difference or a borderline significant difference between groups C and D. This may be due to the small sample size, but we can also see that groups A and B come from vocational schools, while groups C and D come from regular high schools. This suggests that the educational background may also be a factor in this difference between groups. Students in vocational schools are generally older and learn subjects that are closely related to vocational practice. This means that they are more familiar with the knowledge and can integrate new knowledge more quickly. In contrast, students in general high schools are younger and are exposed to video topics that tend to be more academic in nature. This often requires more pattern construction and abstract thinking. This may explain why groups C and D showed more significant differences in cognitive load under VC conditions.

4. Limitations and future directions

Although this study has achieved initial results in exploring the impact of video classrooms on students' cognitive load compared to traditional classrooms, there are still some limitations, which provide a direction for future research to improve. First, this study lacks a direct test of learning outcomes. Although the results of this study found that students in the video classroom had lower HRV during the development phase, we cannot confirm whether this ultimately translates into better learning outcomes. Although previous studies have shown that low HRV is associated with better learning outcomes (Yoo et al., 2021), this result is not definitive. Therefore, future studies should include a test of learning outcomes.

Second, although HRV provides objective biofeedback data, a single measure may not be sufficient to fully capture the sources and effects of cognitive load. Future research could introduce multimodal analysis methods that combine HRV with students' behavioral data (Xue et al., 2024). This could provide a better understanding of students' dynamic cognitive states during complex learning tasks.

Another limitation of this study was the small sample size, within and between groups. Collecting HRV data in classes is logistically demanding. However, this study provides initial indications of the value of analyzing HRV in order to better understand the cognitive load of students, and future studies could draw upon on larger cohorts of students from a wider range of instructional contexts to enhance the generalizability of results.

Finally, despite our efforts to control variables, some factors remained beyond our control. Even though each teacher strictly followed the teaching plan and the content of the two lessons was relevant, some differences still existed. For instance, one lesson might have included more complex concepts. Also, the fact that the study allowed teachers to flexibly schedule the use of video tutorials according to the lesson plan may have also affected the results. For example, students in Group D, who were exposed to the video tutorial only in the second lesson, may have been more familiar with the teaching process, which may have affected their performance. In addition, due to the different lesson scheduling times between groups, there may have been differences in students' learning states and environmental conditions. Future studies can further address this issue through more rigorous experimental controls.

5. Conclusions

Our results show that video tutorials can increase students' cognitive engagement during the knowledge acquisition phase more than traditional teaching materials. This aligns with existing research that emphasizes the advantages of video tutorials in increasing engagement (French et al., 2023; Guo et al., 2014; Lackmann et al., 2021; Zhang et al., 2004). However, our study goes a step further by dynamically tracking how cognitive engagement changes over time. In particular, we found that students mobilized more cognitive resources when using video tutorials and showed higher levels of attention in class. This provides strong evidence that video tutorials, as a form of computer-assisted instruction, are beneficial for learning. Not only do they help to increase students' learning engagement, they also reduce the cognitive load wasted due to external distractions.

The practical implications of this study are manifold. First, video tutorials provide flexible learning support through dynamic demonstrations. Students can control the pace of learning, for example by pausing, rewinding or replaying key content. In addition, our study found that video tutorials reduce distractions among students compared to independent learning or group discussions in traditional classrooms. Finally, video tutorials guide students to invest more cognitive resources in the development phase, which provides empirical evidence for teachers to optimize instructional strategies. In the teaching of complex concepts or core knowledge, prioritizing the use of video tutorials can enhance students' cognitive engagement.

The innovative aspect of our study lies in the application of HRV to monitor cognitive load in a real classroom environment, providing evidence of dynamic cognitive load changes as described in cognitive load theory (Leppink et al., 2013; Paas et al., 2016). Unlike previous biofeedback research conducted mainly in controlled laboratory settings, our study uniquely captured cognitive load fluctuations throughout an entire class session. This real-time insight into how cognitive load shifts across different teaching phases fills a gap in previous research, as most biofeedback-based cognitive load research has been static or limited to experimental settings (Lin et al., 2023).

Overall, our research provides further evidence supporting computer-assisted teaching methods. By dynamically tracking cognitive load using HRV, our study opens new avenues for future research methods, suggesting that educators can use biofeedback technologies to better understand students' cognitive engagement.

Author Contributions

All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Enqi Fan and Jens Siemon. The first draft of the manuscript was written by Enqi Fan. Matt Bower, Jens Siemon revised it critically for important intellectual content. All authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Finding

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

References

Akçapınar, G., Er, E., & Bayazıt, A. (2024). Decoding Video Logs: Unveiling Student Engagement Patterns in Lecture Capture Videos. International Review of Research in Open and Distributed Learning, 25(2), 94-113. https://doi.org/10.19173/irrodl.v25i2.7621

Anders, G., Buder, J., Merkt, M., Egger, E., & Huff, M. (2024). Associations between mind wandering, viewer interactions, and the meaningful structure of educational videos. Computers & Education, 212, 104996. https://doi.org/10.1016/j.compedu.2024.104996

Antonenko, P., Paas, F., Grabner, R., & van Gog, T. (2010). Using Electroencephalography to Measure Cognitive Load. Educational Psychology Review, 22(4), 425–438. https://doi.org/10.1007/s10648-010-9130-y

Arutyunova, K. R., Bakhchina, A. V., Konovalov, D. I., Margaryan, M., Filimonov, A. V., & Shishalov, I. S. (2024). Heart rate dynamics for cognitive load estimation in a driving simulation task. Scientific Reports, 14(1), 31656. https://doi.org/10.1038/s41598-024-79728-x

Author. (forthcoming). Classroom observation coding manual system: Lesson phases, social forms, and time-on-task. Publisher.
Ayres, P. (2006). Using subjective measures to detect variations of intrinsic cognitive load within problems. Learning and Instruction, 16(5), 389–400. https://doi.org/10.1016/j.learninstruc.2006.09.001
Ayres, P. (2017). Subjective Measures of Cognitive Load. Cognitive Load Measurement and Application, 9–28. https://doi.org/10.4324/9781315296258-2
Baceviciute, S., Lucas, G., Terkildsen, T., & Makransky, G. (2022). Investigating the redundancy principle in immersive virtual reality environments: An eye-tracking and EEG study. Journal of Computer Assisted Learning, 38(1), 120-136. https://doi.org/10.1111/jcal.12595
Berka, C., Levendowski, D. J., Lumicao, M. N., Yau, A., Davis, G., Zivkovic, V. T., Olmstead, R. E., Tremoulet, P. D., & Craven, P. L. (2007). EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks. Aviation, space, and environmental medicine, 78(5), B231-B244.
Bernardi, L., Wdowczyk-Szulc, J., Valenti, C., Castoldi, S., Passino, C., Spadacini, G., & Sleight, P. (2000). Effects of controlled breathing, mental activity and mental stress with or without verbalization on heart rate variability. Journal of the American College of Cardiology, 35(6), 1462–1469. https://doi.org/10.1016/s0735-1097(00)00595-7
Biard, N., Cojean, S., & Jamet, E. (2018). Effects of segmentation and pacing on procedural learning by video. Computers in Human Behavior, 89, 411–417. https://doi.org/10.1016/j.chb.2017.12.002
Brunken, R., Plass, J. L., & Leutner, D. (2003). Direct Measurement of Cognitive Load in Multimedia Learning. Educational Psychologist, 38(1), 53–61. https://doi.org/10.1207/s15326985ep3801_7
Bueno-Vesga, J. A., Xu, X., & He, H. (2021). The effects of cognitive load on engagement in a virtual reality learning environment. 2021 IEEE Virtual Reality and 3D User Interfaces (VR). https://doi.org/10.1109/vr50410.2021.00090
Champseix, R., Ribiere, L., & Le Couedic, C. (2021). A python package for heart rate variability analysis and signal preprocessing. Journal of Open Research Software, 9(1), 28. https://doi.org/10.5334/jors.305
Chen, C. M., & Wu, C. H. (2015). Effects of different video lecture types on sustained attention, emotion, cognitive load, and learning performance. Computers & Education, 80, 108–121. https://doi.org/10.1016/j.compedu.2014.08.015
Chen, H. T. M., & Thomas, M. (2020). Effects of lecture video styles on engagement and learning. Educational Technology Research and Development, 68, 2147-2164. https://doi.org/10.1007/s11423-020-09757-6
Chen, S., Epps, J., & Chen, F. (2011). A comparison of four methods for cognitive load measurement. Proceedings of the 23rd Australian Computer-Human Interaction Conference. https://doi.org/10.1145/2071536.2071547
Cierniak, G., Scheiter, K., & Gerjets, P. (2009). Explaining the split-attention effect: Is the reduction of extraneous cognitive load accompanied by an increase in germane cognitive load? Computers in Human Behavior, 25(2), 315–324. https://doi.org/10.1016/j.chb.2008.12.020
Conway, D., Dick, I., Li, Z., Wang, Y., & Chen, F. (2013). The Effect of Stress on Cognitive Load Measurement. Human-Computer Interaction – INTERACT 2013, 659–666. https://doi.org/10.1007/978-3-642-40498-6_58
Costley, J., & Lange, C. H. (2017). The effects of lecture diversity on Germane Load. The International Review of Research in Open and Distributed Learning, 18(2). https://doi.org/10.19173/irrodl.v18i2.2860
Costley, J., Fanguy, M., Lange, C., & Baldwin, M. (2021). The effects of video lecture viewing strategies on cognitive load. Journal of Computing in Higher Education, 33, 19-38. https://doi.org/10.1007/s12528-020-09254-y
Cranford, K. N., Tiettmeyer, J. M., Chuprinko, B. C., Jordan, S., & Grove, N. P. (2014). Measuring Load on Working Memory: The Use of Heart Rate as a Means of Measuring Chemistry Students' Cognitive Load. Journal of Chemical Education, 91(5), 641–647. https://doi.org/10.1021/ed400576n
de Jong, T. (2009). Cognitive load theory, educational research, and instructional design: some food for thought. Instructional Science, 38(2), 105–134. https://doi.org/10.1007/s11251-009-9110-0
DeLeeuw, K. E., & Mayer, R. E. (2008). A comparison of three measures of cognitive load: Evidence for separable measures of intrinsic, extraneous, and germane load. Journal of Educational Psychology, 100(1), 223–234. https://doi.org/10.1037/0022-0663.100.1.223
Deng, R. (2024). Effect of video styles on learner engagement in MOOCs. Technology, Pedagogy and Education, 33(1), 1-21. https://doi.org/10.1080/1475939X.2023.2246981
Ding, L., Yoon, M., & Kim, D. (2024). Unveiling the pedagogical advantage of tutoring-style videos in an authentic biology class. Journal of Computer Assisted Learning, 40(3), 946-959. https://doi.org/10.1111/jcal.12930
Dönmez, O., Akbulut, Y., Telli, E., Kaptan, M., Özdemir, İ. H., & Erdem, M. (2022). In search of a measure to address different sources of cognitive load in computer-based learning environments. Education and Information Technologies, 27(7), 10013-10034. https://doi.org/10.1007/s10639-022-11035-2
Egger, M., Ley, M., & Hanke, S. (2019). Emotion recognition from Physiological Signal Analysis: A Review. Electronic Notes in Theoretical Computer Science, 343, 35–55. https://doi.org/10.1016/j.entcs.2019.04.009
Fan, E., Bower, M., & Siemon, J. (2024). Video Tutorials in the Traditional Classroom: The Effects on Different Types of Cognitive Load. Technology, Knowledge and Learning, 1-20. https://doi.org/10.1007/s10758-024-09754-1
Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. Behavior Research Methods, 41(4), 1149–1160. https://doi.org/10.3758/brm.41.4.1149
French, S., Ravn, S., Balcaite, E., & Moore, E. (2023). Understanding students' views on the efficacy of video technology to promote engagement in higher education. Pacific Journal of Technology Enhanced Learning, 5(2), 1-14. https ://doi.org/10.24135/pjtel.v1i2.172

Gilgen-Ammann, R., Schweizer, T., & Wyss, T. (2019). RR interval signal quality of a heart rate monitor and an ECG Holter at rest and during exercise. European Journal of Applied Physiology, 119(7), 1525–1532. https://doi.org/10.1007/s00421-019-04142-5

Grassmann, M., Vlemincx, E., von Leupoldt, A., & Van den Bergh, O. (2015). The role of respiratory measures to assess mental load in pilot selection. Ergonomics, 59(6), 745–753. https://doi.org/10.1080/00140139.2015.1090019

Guo, P. J., Kim, J., & Rubin, R. (2014, March). How video production affects student engagement: An empirical study of MOOC videos. In Proceedings of the first ACM conference on Learning@ scale conference (pp. 41-50). https://doi.org/10.1145/2556325.2566239

Haapalainen, E., Kim, S., Forlizzi, J. F., & Dey, A. K. (2010). Psycho-physiological measures for assessing cognitive load. Proceedings of the 12th ACM International Conference on Ubiquitous Computing. https://doi.org/10.1145/1864349.1864395

Hjortskov, N., Rissén, D., Blangsted, A. K., Fallentin, N., Lundberg, U., & Søgaard, K. (2004). The effect of mental stress on heart rate variability and blood pressure during computer work. European Journal of Applied Physiology, 92(1–2), 84–89. https://doi.org/10.1007/s00421-004-1055-z

Hong, J., Pi, Z., & Yang, J. (2016). Learning declarative and procedural knowledge via video lectures: Cognitive Load and learning effectiveness. Innovations in Education and Teaching International, 55(1), 74–81. https://doi.org/10.1080/14703297.2016.1237371

Hughes, C., Costley, J., & Lange, C. (2018). The effects of self-regulated learning and cognitive load on beginning to watch and completing video lectures at a cyber-university. Interactive Technology and Smart Education, 15(3), 220–237. https://doi.org/10.1108/itse-03-2018-0018

Kalyuga, S. (2011). Cognitive load theory: How many types of load does it really need?. Educational psychology review, 23, 1-19. https://doi.org/10.1007/s10648-010-9150-7

Karch, J. M., García Valles, J. C., & Sevian, H. (2019). Looking into the Black Box: Using Gaze and Pupillometric Data to Probe How Cognitive Load Changes with Mental Tasks. Journal of Chemical Education, 96(5), 830–840. https://doi.org/10.1021/acs.jchemed.9b00014

Kirschner, P. A. (2002). Cognitive load theory: implications of cognitive load theory on the design of learning. Learning and Instruction, 12(1), 1–10. https://doi.org/10.1016/s0959-4752(01)00014-7

Klepsch, M., Schmitz, F., & Seufert, T. (2017). Development and Validation of Two Instruments Measuring Intrinsic, Extraneous, and Germane Cognitive Load. Frontiers in Psychology, 8. https://doi.org/10.3389/fpsyg.2017.01997

Krieglstein, F., Beege, M., Rey, G.D. et al. Development and Validation of a Theory-Based Questionnaire to Measure Different Types of Cognitive Load. Educ Psychol Rev 35, 9 (2023). https://doi.org/10.1007/s10648-023-09738-0

Kuhlmann, S. L., Plumley, R., Evans, Z., Bernacki, M. L., Greene, J. A., Hogan, K. A., ... & Panter, A. (2024). Students' active cognitive engagement with instructional videos predicts STEM learning. Computers & Education, 216, 105050. https://doi.org/10.1016/j.compedu.2024.105050

Lackmann, S., Léger, P. M., Charland, P., Aubé, C., & Talbot, J. (2021). The influence of video format on engagement and performance in online learning. Brain Sciences, 11(2), 128. https://doi.org/10.3390/brainsci11020128

Lan, C.-H., Sheng, M.-H., Hsu, Y.-C., & Shiue, Y.-M. (2019). Influence of online and face-to-face collaboration and learning style on cognitive load and engagement in a health introductory course. Proceedings of the Third International Conference on Medical and Health Informatics 2019 - ICMHI 2019. https://doi.org/10.1145/3340037.3340069

Larmuseau, C., Vanneste, P., Cornelis, J., Desmet, P., & Depaepe, F. (2019). Combining physiological data and subjective measurements to investigate cognitive load during complex learning. Frontline Learning Research, 7(2), 57–74. https://doi.org/10.14786/flr.v7i2.403

Leppink, J., Paas, F., Van der Vleuten, C. P., Van Gog, T., & Van Merriënboer, J. J. (2013). Development of an instrument for measuring different types of cognitive load. Behavior Research Methods, 45(4), 1058–1072. https://doi.org/10.3758/s13428-013-0334-1

Levenson, R. W. (2003). Blood, sweat, and fears. Annals of the New York Academy of Sciences, 1000(1), 348–366. https://doi.org/10.1196/annals.1280.016

Liao, C. H., & Wu, J. Y. (2023). Learning analytics on video-viewing engagement in a flipped statistics course: Relating external video-viewing patterns to internal motivational dynamics and performance. Computers & Education, 197, 104754. https://doi.org/10.1016/j.compedu.2023.104754

Lin, X., Tang, W., Ma, W., Liu, Y., & Ding, F. (2023). The impact of media diversity and cognitive style on learning experience in programming video lecture: A brainwave analysis. Education and Information Technologies, 28(8), 10617-10637. https://doi.org/10.1007/s10639-023-11608-9

Lloyd, S. A., & Robertson, C. L. (2011). Screencast Tutorials Enhance Student Learning of Statistics. Teaching of Psychology, 39(1), 67–71. https://doi.org/10.1177/0098628311430640

Luke, B., & Hogarth, K. (2011). Developing and enhancing independent learning skills. Accounting Research Journal, 24(3), 290–310. https://doi.org/10.1108/10309611111187019

Makransky, G., Terkildsen, T. S., & Mayer, R. E. (2019). Adding immersive virtual reality to a science lab simulation causes more presence but less learning. Learning and instruction, 60, 225-236. https://doi.org/10.1016/j.learninstruc.2017.12.007

Mayer, R. E. (2005). Cognitive Theory of Multimedia Learning. The Cambridge Handbook of Multimedia Learning, 31–48. https://doi.org/10.1017/cbo9780511816819.004

McDuff, D., Gontarek, S., & Picard, R. (2014). Remote measurement of cognitive stress via heart rate variability. 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. https://doi.org/10.1109/embc.2014.6944243

McHugh, M. L. (2012). Interrater Reliability: The kappa statistic. Biochemia Medica, 276–282. https://doi.org/10.11613/bm.2012.031

Miller, B. W. (2015). Using reading times and eye-movements to measure cognitive engagement. Educational Psychologist, 50(1), 31–42. doi:10.1080/00461520.2015.100406

Minkley, N., Kärner, T., Jojart, A., Nobbe, L., & Krell, M. (2018). Students' mental load, stress, and performance when working with symbolic or symbolic–textual molecular representations. Journal of Research in Science Teaching, 55(8), 1162–1187. https://doi.org/10.1002/tea.21446

Mizuno, K., Tanaka, M., Yamaguti, K., Kajimoto, O., Kuratsune, H., & Watanabe, Y. (2011). Mental fatigue caused by prolonged cognitive load associated with sympathetic hyperactivity. Behavioral and Brain Functions, 7(1), 17. https://doi.org/10.1186/1744-9081-7-17

Moya-Ramon, M., Mateo-March, M., Peña-González, I., Zabala, M., & Javaloyes, A. (2022). Validity and reliability of different smartphones applications to measure HRV during short and ultra-short measurements in elite athletes. Computer Methods and Programs in Biomedicine, 217, 106696. https://doi.org/10.1016/j.cmpb.2022.106696

Mukherjee, S., Yadav, R., Yung, I., Zajdel, D. P., & Oken, B. S. (2011). Sensitivity to mental effort and test-retest reliability of heart rate variability measures in healthy seniors. Clinical Neurophysiology. https://doi.org/10.1016/j.clinph.2011.02.032

Mullikin, D. R., Flanagan, R. P., Merkebu, J., Durning, S. J., & Soh, M. (2024). Physiologic measurements of cognitive load in clinical reasoning. Diagnosis, 11(2), 125-131. https://doi.org/10.1515/dx-2023-0143

Murphy, D. H., Hoover, K. M., Agadzhanyan, K., Kuehn, J. C., & Castel, A. D. (2022). Learning in double time: The effect of lecture video speed on immediate and delayed comprehension. Applied Cognitive Psychology, 36(1), 69-82. https://doi.org/10.1002/acp.3899

Örün, Ö., & Akbulut, Y. (2019). Effect of multitasking, physical environment and electroencephalography use on cognitive load and retention. Computers in Human Behavior, 92, 216-229. https://doi.org/10.1016/j.chb.2018.11.027

Ouwehand, K., Kroef, A. V. D., Wong, J., & Paas, F. (2021, September). Measuring cognitive load: Are there more valid alternatives to likert rating scales?. In Frontiers in Education (Vol. 6, p. 702616). Frontiers Media SA. https://doi.org/10.3389/feduc.2021.702616

Paas, F. G. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. Journal of Educational Psychology, 84(4), 429–434. https://doi.org/10.1037/0022-0663.84.4.429

Paas, F. G. W. C., Ayres, P., & Pachman, M. (2008). Assessment of cognitive load in multimedia learning. Recent Innovations in Educational Technology That Facilitate Student Learning, Information Age Publishing Inc., Charlotte, NC, 11-35.

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive Load Theory and Instructional Design: Recent Developments. Educational Psychologist, 38(1), 1–4. https://doi.org/10.1207/s15326985ep3801_1

Paas, F., Renkl, A., & Sweller, J. (Eds.). (2016). Cognitive load theory: A Special issue of educational psychologist. Routledge.

Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. (2003). Cognitive Load Measurement as a Means to Advance Cognitive Load Theory. Educational Psychologist, 38(1), 63–71. https://doi.org/10.1207/s15326985ep3801_8

Parker, M. J., Bunch, M., & Pike, A. (2024). Is anybody watching: A multi-factor motivational framework for educational video engagement. Computers & Education, 105148. https://doi.org/10.1016/j.compedu.2024.105148

Pham, T., Lau, Z. J., Chen, S. H., & Makowski, D. (2021). Heart rate variability in psychology: A review of HRV Indices and an analysis tutorial. Sensors, 21(12), 3998. https://doi.org/10.3390/s21123998

Plass, J. L., Kalyuga, S., & Leutner, D. (2010). Individual Differences and Cognitive Load Theory. Cognitive Load Theory, 65–88. https://doi.org/10.1017/cbo9780511844744.006

Rolim, L. C., de Souza, J. S., & Dib, S. A. (2013). Tests for early diagnosis of cardiovascular autonomic neuropathy: Critical analysis and relevance. Frontiers in Endocrinology, 4. https://doi.org/10.3389/fendo.2013.00173

Rumelhart, D. E. (1980). On Evaluating Story Grammars*. Cognitive Science, 4(3), 313–316. https://doi.org/10.1207/s15516709cog0403_5

Sharma, R. (2017). Computer assisted learning–A study. International Journal of Advanced Research in Education & Technology (IJARET), 4(2), 102-105.

Šola, H. M., Qureshi, F. H., & Khawaja, S. (2024). AI Eye-Tracking Technology: A New Era in Managing Cognitive Loads for Online Learners. Education Sciences, 14(9), 933. https ://doi.org/10.3390/educsci14090933

Solhjoo, S., Haigney, M. C., McBee, E., van Merrienboer, J. J., Schuwirth, L., Artino, A. R., Battista, A., Ratcliffe, T. A., Lee, H. D., & Durning, S. J. (2019). Heart rate and heart rate variability correlate with clinical reasoning performance and self-reported measures of cognitive load. Scientific Reports, 9(1). https://doi.org/10.1038/s41598-019-50280-3

Song, H. S., & Lehrer, P. M. (2003). The effects of specific respiratory rates on heart rate and heart rate variability. Applied psychophysiology and biofeedback, 28, 13-23. https://doi.org/10.1023/A:1022312815649

Souchet, A. D., Philippe, S., Lourdeaux, D., & Leroy, L. (2022). Measuring visual fatigue and cognitive load via eye tracking while learning with virtual reality head-mounted displays: A review. International Journal of Human–Computer Interaction, 38(9), 801-824. https://doi.org/10.1080/10447318.2021.1976509

Speer, K. E., Semple, S., Naumovski, N., & McKune, A. J. (2020). Measuring heart rate variability using commercially available devices in healthy children: A validity and Reliability Study. European Journal of Investigation in Health, Psychology and Education, 10(1), 390–404. https://doi.org/10.3390/ejihpe10010029

Stein, P. K., Bosner, M. S., Kleiger, R. E., & Conger, B. M. (1994). Heart rate variability: A measure of cardiac autonomic tone. American Heart Journal, 127(5), 1376–1381. https://doi.org/10.1016/0002-8703(94)90059-0

Sweller, J. (1988). Cognitive Load During Problem Solving: Effects on Learning. Cognitive Science, 12(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4

Sweller, J. (2023). Cognitive load theory: What we learn and how we learn. In Learning, design, and technology: An international compendium of theory, research, practice, and policy (pp. 137-152). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-17461-7_50

Sweller, J. Element Interactivity and Intrinsic, Extraneous, and Germane Cognitive Load. Educ Psychol Rev 22, 123–138 (2010). https://doi.org/10.1007/s10648-010-9128-5

Sweller, J., van Merrienboer, J. J., & Paas, F. G. (1998). Cognitive architecture and instructional design. Educational Psychology Review, 10(3), 251–296. https://doi.org/10.1023/a:1022193728205

Thong, T., Li, K., McNames, J., Aboy, M., & Goldstein, B. (2003). Accuracy of ultra-short heart rate variability measures. Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEEE Cat. No.03CH37439). https://doi.org/10.1109/iembs.2003.1280405

Tseng, S. S. (2021). The influence of teacher annotations on student learning engagement and video watching behaviors. International Journal of Educational Technology in Higher Education, 18(1), 7. https://doi.org/10.1186/s41239-021-00242-5

van der Meij, H., & van der Meij, J. (2014). A comparison of paper-based and video tutorials for software learning. Computers & Education, 78, 150–159. https://doi.org/10.1016/j.compedu.2014.06.003

Wahyudi, S., Joyoatmojo, S., & Sawiji, H. (2017). Learning Model of Attention, Relevance, Confidence, Satisfaction (ARCS) Supported by Video Tutorial to Improve the Students' Learning Motivation in Vocational High School. Proceedings of the International Conference on Teacher Training and Education 2017 (ICTTE 2017). https://doi.org/10.2991/ictte-17.2017.72

Wells, J., Barry, R. M., & Spence, A. (2012). Using Video Tutorials as a Carrot-and-Stick Approach to Learning. IEEE Transactions on Education, 55(4), 453–458. https://doi.org/10.1109/te.2012.2187451

Whelan, R. R. (2007). Neuroimaging of cognitive load in instructional multimedia. Educational Research Review, 2(1), 1–12. https://doi.org/10.1016/j.edurev.2006.11.001

Wu, S. F., Kao, C. H., Lu, Y. L., & Lien, C. J. (2022). A Method Detecting Student's Flow Construct during School Tests through Electroencephalograms (EEGs): Factors of Cognitive Load, Self-Efficacy, Difficulty, and Performance. Applied Sciences, 12(23), 12248. https://doi.org/10.3390/app122312248

Xie, B., & Salvendy, G. (2000). Review and reappraisal of modelling and predicting mental workload in single-and multi-task environments. Work & stress, 14(1), 74-99. https://doi.org/10.1080/026783700417249

Xue, Y., Wang, K., & Qiu, Y. (2024). Enhancing Online Learning: A Multimodal Approach for Cognitive Load Assessment. International Journal of Human–Computer Interaction, 1-11. https://doi.org/10.1080/10447318.2024.2327198

Yoo, H. H., Yune, S. J., Im, S. J., Kam, B. S., & Lee, S. Y. (2021). Heart rate variability-measured stress and academic achievement in medical students. Medical Principles and Practice, 30(2), 193-200. https://doi.org/10.1159/000513781

Zhang, D., Zhao, J. L., Zhou, L., & Nunamaker Jr, J. F. (2004). Can e-learning replace classroom learning?. Communications of the ACM, 47(5), 75-79. https://doi.org/10.1145/986213.986216

Zheng, X., Ma, Y., Yue, T., & Yang, X. (2023). Effects of different types of cues and self-explanation prompts in instructional videos on Deep learning: Evidence from multiple data analysis. Educational Technology Research and Development, 71(3), 807–831. https://doi.org/10.1007/s11423-023-10188-2

Zhou, W. G., Yu, P. P., Wu, L. H., Cao, Y. F., Zhou, Y., & Yuan, J. J. (2024). Pilot turning behavior cognitive load analysis in simulated flight. Frontiers in Neuroscience, 18, 1450416. https://doi.org/10.3389/fnins.2024.1450416

Ziemssen, T., & Siepmann, T. (2019). The investigation of the cardiovascular and sudomotor autonomic nervous system—a review. Frontiers in neurology, 10, 53. https://doi.org/10.3389/fneur.2019.00053

Appendix A.

A.1. Classroom Observation Coding Manual System- Lesson Phases, Social Forms, and Time-on-Task

Learning and Instruction

From Heartbeats to Actions:

Multimodal Learning Analytics of Cognitive and Behavior Engagement in Real Classrooms

Enqi Fan[a*], Matt Bower[b], Jens Siemon[a]

[a]*Department of Vocational and Business Education, University of Hamburg, Sedanstraße 19, 20146, Hamburg, Germany*
[b]*School of Education, Macquarie University, Building 29WW, Balaclava Rd North Ryde, 2109, NSW Sydney, Australia*

Structured Abstract

Background: In the classroom, there may be a mismatch between students' external behavior and their internal cognitive states. For example, students may exhibit attentive behavior but not be engaged in deep cognitive processing.
Aims: This study combined time-on-task and heart rate variability (HRV) data to explore the dynamic relationship between student behavior and cognition. By comparing Video Learning Classrooms (VLC) with Traditional Learning Classrooms (TLC), the influence of different teaching settings on students' learning states was explored.
Sample: 45 students (paired sample) from German vocational schools and high schools.
Methods: Time-on-task in the classroom was coded (in 10-second intervals) as a measure of students' behavior. The HRV indicator RMSSD (root mean square of successive differences) measured students' cognition.
Results: Kruskal-Wallis tests revealed significant differences in HRV between the two classroom settings for all time-on-task categories ($p < 0.001$). Spearman's correlation analysis showed a significant negative correlation between time-on-task and HRV (VLC: $\rho = -0.1621$; TLC: $\rho = -0.2184$). In the VLC, students had more cognitive load when performing learning tasks, but the cognitive fluctuations were more stable. The TLC, on the other hand, showed greater cognitive fluctuations
Conclusion: Combining behavioral and physiological data using the MMLA can more accurately capture students' learning states. The study found that the impact of different instructional settings on students' learning states differs. Video tutorials help to stabilize students' cognitive engagement, while traditional classrooms place higher demands on students' self-regulation abilities.

Keywords: multimodal learning analysis, heart rate variability, cognitive engagement, behavior engagement, time-on-task

* Corresponding author. *E-mail address:* enqi.fan@uni-hamburg.de
Submitted: October 2024

1. Introduction

A situation may arise when students are listening to the teacher or studying on their own: they may appear to be fully engaged, but deep cognitive processing may not be taking place in their brains. This inconsistency poses a challenge for teachers: how can they tell if students are truly engaged in the learning process? The complexity of this issue comes from the dynamic nature of cognition. Students may appear to be engaged in classroom activities, but their cognitive engagements may be quite different. This difference between behavior and cognition can affect instructional decisions.

Classroom video analysis can provide some assistance to teachers (Klette, 2023). By observing classroom videos, teachers can check students' behavior such as interactions and performances in different teaching sessions (Buijs & Admiraal, 2013; Ghergulescu & Muntean, 2016; Spanjers et al., 2008). Time-on-task is widely used as an important indicator of students' learning engagement in classroom behavior analysis. It refers to the amount of time that students are actively devoted to learning while completing learning tasks (Hesse, 1994). It has been one of the most crucial variables for measuring classroom performance and learning outcomes since the 1970s (Carvalho et al., 2017; Fisher, 1978; Fredrick et al., 1979; Scholkmann et al., 2017). However, as a measure of external behavior, can time-on-task also reveal students' deeper states at the cognitive level? For example, do students work as hard at the cognitive level when they are in high time-on-task?

Cognitive load theory (CLT; Sweller, 1988) provides a theoretical framework for examining this issue. CLT argues that the allocation of students' cognitive resources is a key factor in the learning process. When learners devote more cognitive resources, their engagement is higher (Miller, 2015). CLT distinguishes between intrinsic load (inherent task complexity), extraneous load (unnecessary cognitive demands due to poor instructional design), and germane load (effort devoted to schema construction during learning; Kalyuga, 2011; Sweller et al., 1998). It is worth noting that CLT emphasizes the role of interactive elements in task complexity, especially in learning tasks. However, in non-learning phases of the classroom, such as organizing activities, task complexity may come from different sources (Ayres et al., 2021). This highlights the need for a broader framework that considers not only the cognitive load in learning tasks, but also cognitive engagement in other lesson phases. Cognitive engagement refers to the psychological investment that students put into the learning process (Corno & Mandinach, 1983). It emphasizes the mental effort of thinking and applying strategies. It also includes the willingness to understand complex concepts to cope with challenging material and acquire new knowledge (Fredricks et al., 2004). Cognitive engagement complements cognitive load by focusing on students' psychological investment. It is not limited to learning tasks, but also covers the broader classroom phase.

However, measuring cognitive load and cognitive engagement is a complex process, especially in real classrooms. Traditional measurement methods such as questionnaires, although easy to use, are subjective and limit the capture of students' real-time cognitive states. In recent years, with the advancement of technology, biofeedback technology has provided a new mechanism for exploring cognitive load. For example, electroencephalography (EEG) and functional magnetic resonance imaging (fMRI; Antonenko et al., 2010; Hassib et al., 2017; Vitolo et al., 2021). However, due to their high cost and strict experimental conditions, they are difficult to apply in real classrooms. In contrast, heart rate variability (HRV) is a physiological indicator that reflects autonomic nervous system (ANS) activity through changes in heartbeat intervals. Due to its portability, low cost, and non-invasiveness, it is an ideal tool for studying cognitive dynamics in the classroom (Ziemssen

& Siepmann, 2019; Elkin et al., 2024). Some studies have shown that when cognitive load increases, HRV decreases. Conversely, when cognitive load decreases, HRV increases (Mukherjee et al., 2011; Pham et al., 2021). Therefore, HRV can provide an objective measure of the dynamic changes in cognitive level.

This study proposes a multimodal learning analysis framework for analyzing students' cognitive load and engagement by combining HRV and Time-on-task data. Specifically, we synchronize HRV and Time-on-task data to expose the relationship between students' external learning behavior and the dynamic changes in their cognitive engagement. Compared with viewing time-on-task or cognitive engagement alone, this approach more comprehensively shows the relationship between students' behaviors and cognition during the learning process. We conducted experiments in two teaching environments: the Video Learning Classroom (VLC) and the Traditional Learning Classroom (TLC), to compare the effects of different instructional methods on students' cognitive and behavior engagement. The results provide evidence which can be used to support the optimization of instructional design and intervention strategies.

1.1. Classroom observation and time on task

Classroom observation is an important educational research method (Klette, 2015; Dignath & Veenman, 2020). By observing the behavior of students and teachers in the classroom, researchers can optimize the assessment of teaching quality and teaching strategies in a way that is grounded in the reality of actual classes (Granström et al., 2023; Klette, 2023). Video analysis, as an advanced form of classroom observation, can clearly record classroom behaviors (Fischer & Neumann, 2012). It enables researchers to combine qualitative and quantitative analysis to reveal how changes in teaching activities impact on students through observation and coding (Borko et al., 2008; Klette, 2023).

Time-on-task is an important indicator in classroom observation. It refers to the time that students actively spend on learning in order to complete learning tasks (Anderson, 1995; Hesse, 1994). The "time-on-task hypothesis" argues that the more time students devote to learning, the better the learning outcomes and the more effective the teacher's instructional design (Carroll, 1963; Gettinger & Seibert, 2002; Helmke, 2007; Lipowsky, 2006). Some studies use Time-on-task to measure students' classroom performance (Götzl et al., 2013; Knigge et al., 2013; Siemon et al., 2015; Scholkmann et al., 2017) and classroom participation (Buijs & Admiraal, 2013; Ghergulescu & Muntean, 2016; Spanjers et al., 2008).

1.2. Objective Measurement of Cognitive load and Cognitive engagement

For a long time, subjective measurement has been the main way to measure cognitive load and cognitive engagement (Ayres, 2006; Greene, 2015; Paas, 1992). Besides traditional subjective measurement methods, objective measurement methods that collect biofeedback data have also gradually been used (Li, 2021; Xiong et al.,2020). HRV can reflect the dynamic changes of the learner's autonomic nervous system (ANS), which are manifested as small fluctuations in heartbeat intervals (Ziemssen & Siepmann, 2019). The ANS changes when learners are engaged in class (Levenson, 2003). The ANS is mainly composed of the sympathetic nervous system (SNS) and the parasympathetic nervous system (PNS), which take turns in different states (Egger et al., 2019). Under stress conditions, the SNS is more active, resulting in an increase in heart rate and a decrease in HRV. In a relaxed state, the PNS dominates, which results in a lower heart rate and an increased HRV (Pham et al., 2021). HRV is therefore considered a reliable physiological indicator for assessing the state and response of the ANS (Bernardi et al., 2000; Hjortskov et al., 2004; Mullikin et al., 2024; Rolim et al., 2013).

In addition, studies have shown that changes in HRV are closely related to cognitive load. When cognitive load increases, it can cause significant changes in physiological indicators, including faster breathing and higher blood pressure, which can reduce HRV levels (Grassmann et al., 2015; Song & Lehrer, 2003; Solhjoo et al., 2019). This characteristic makes HRV an effective means of monitoring cognitive level, and it is widely used in cognitive science and educational research (Forte et al., 2019). By measuring HRV, researchers can gain a deeper understanding of cognitive changes during the learning process (Haapalainen et al., 2010; Mizuno et al., 2011; Minkley et al., 2018; Mukherjee et al., 2011).

1.3. The present study

This study will explore the dynamic relationship between student behavior engagement and cognitive engagement through time-on-task and HRV data. Multimodal Learning Analytics (MMLA) offers new possibilities for integrating the two types of data (Cowling & Birt, 2020; Giannakos et al., 2022). However, the real classroom environment is more complex than the laboratory, which makes conducting MMLA more challenging (Cukurova et al., 2020; Ouhaichi et al., 2023). Students' behavior and cognitive state in the classroom may be affected by a variety of factors. These factors may include the lesson phase, how the content is presented, environmental distractions, and peer interactions.

To address these challenges, video tutorials were introduced as a classroom intervention in this study. Video tutorials are considered to be a potentially powerful educational tool and have a possible impact on cognitive load (Hong et al., 2016; Paas et al., 2008; Sharma, 2017). By comparing video-learning classrooms (VLC) with traditional learning classrooms (TLC), this study explored how different instructional designs affect students' behavior and cognitive engagement.

This study proposes the following research questions:

RQ1: Is there a difference in HRV changes for different types of tasks in the VLC and the TLC?

RQ2: Is there a correlation between HRV changes and Time-on-task changes for different types of tasks in the VLC and the TLC?

This study aims to explore the relationship between students' external learning behaviors and dynamic cognitive changes by integrating Time-on-task and HRV data. Additionally, the study aims to verify the feasibility and applicability of Multimodal Learning Analytics (MMLA) in complex classroom situations and to offer empirical support for its implementation in real instructional environments.

2. Method

2.1. Curriculum and Participants

Four groups of lessons (eight lessons) were recorded for this study. Two of these groups were from two different types of vocational schools in Germany. The other two groups were from a high school in Germany. Looking at different types of teaching environments aimed to make the research results more generalizable. Each group of recordings included two different lessons. The Video Learning Classroom (VLC) used video tutorials when learning new knowledge. The Traditional Learning Classroom (TLC) used traditional teaching methods. In the VLC, the teacher selected the appropriate video tutorials according to the teaching progress and content. All recordings followed the students' original learning plan, which ensured the continuity of the teaching content.

Paired designs were used in this study to control for the influence of individual differences on the results. Only students who participated in both teaching methods were included in the sample. To ensure that the study had sufficient statistical power, the sample size was estimated based on a medium effect size of Cohen's d (0.50). At a 90% confidence level (α = 0.05, two-tailed), the study required at least 44 participants according to G*Power 3.1 calculations (Faul et al., 2009). A total of 45 students (21 males; Mage = 19.75; SDage = 3.67) participated in the study. The specific course settings and numbers are shown in Table 1. The study was approved by the Ethics Committee. All participants were informed about the content of the study in advance and signed an informed consent form. The informed consent forms of underage students were signed by their parents. All data were collected anonymously.

Table. 1. *Group Information and Participant Demographics*

Group	School Type	Topic	$N_{students}$	N_{male}	M_{age}	SD_{age}
A	Vocational school	Security Management	10	7	21.6	4.9
B	Vocational school	Unemployment	9	1	23	3.94
C	High school	Biology	14	8	18	0.41
D	High school	Economics and Politics	12	5	17.83	0.83

2.2. Instruments

In this study, we chose the Polar H10 heart rate band (Polar Electro Oy, Kempele, Finland) as the HRV acquisition device. It has been shown to correctly measure the RR interval (the time between successive R-wave peaks in the heartbeat, reflecting heart rhythm and variability; Gilgen-Ammann et al., 2019; Speer et al., 2020; Moya-Ramon et al., 2022). The Polar H10 heart rate band sampled at a 1000 Hz frequency, and the data was transmitted via Bluetooth connection to four iMotions laptops (9.3, iMotions, Copenhagen, Denmark). All data was saved locally.

2.3. Procedure

Before the recording began, researchers set up four high-definition cameras in the four corners of the classroom. Each student had a voice recorder, a heart rate band, and a numbered card on their desk. Students were given headphones during the VLC.

2.4. Video and Audio Data Preprocessing

Video analysis software Mangold Interact (Mangold International GmbH, Germany) was used to code the classroom videos for this study. Coding was based on two manuals: the Coding Manual - Time-on-Task and the Coding Manual - Lesson Phases (Author, 2024; see Appendix A). The coding manual – Time-on-task samples students' time allocation at 10-second intervals. It includes categories such as private time, organizational time, and real learning time. The coding manual – Lesson phases samples events to mark different parts of the classroom teaching. These include, for example, learning organization, introduction, development, practicing and debriefing. Only a brief description of the coding themes and content of the "Coding Manual - Time-on-Task" is provided here. Please refer to the appendix for the "Coding Manual - Lesson Phases" and specific coding rules and details.

Each coding task was completed independently by two trained coders. After coding is completed, the inter-rater agreement is tested using Cohen's Kappa to ensure the reliability of the results. The Kappa values for the Coding Manual - Time-on-Task range from 0.78 to 0.90. The Kappa values for the Coding Manual - Lesson Phases range from 0.71 to 1.00 (Author, 2024). Coding is only considered complete when the Kappa value meets the manual's specified agreement standard.

Coding Manual - Time-on-Task (time-sample, 10 second units, Cohens Kappa: 0,78-0,90)
Code 1. Private - but not yet finished with the learning task or task processing
Code 2. Organization
Code 3. Real learning time
Code 4. Private - but finished with the learning task or task processing
Code 0. not assignable

This study coded the time-on-task of each participant and generated raw data with a code every 10 seconds (Time-on-task-10s). We then calculated the percentage of time spent on the task per minute (Time-on-task%-60s). To ensure that the percentage of time spent on the task accurately reflected the time students actually spent learning, only codes 1 and 3 were included in the calculation. The formula is: [code3 / (code1 + code3)]. This allowed us to observe the time students actually spent on the learning task.

2.5. Heart Rate Variability Data Preprocessing

To calculate HRV, we exported data from the iMotions notebook containing audio and video timestamps. We cleaned and filtered the data using the 3sigma function (a method for removing outliers). We then applied the root mean square of successive differences (RMSSD) calculation to the RR intervals (Stein et al., 1994). RMSSD can be calculated even with ultra-short heart rate recordings (Thong et al., 2003; Pham et al., 2021). As an indicator of the parasympathetic nervous system, a decrease in RMSSD values generally indicates an increase in cognitive load and engagement.

In this study, we calculated the RMSSD every 10 seconds (RMSSD-10s) and every 60 seconds (RMSSD-60s) for each student to synchronize the HRV data with the Time-on-task-10s and Time-on-task%-60s data. Using RMSSD-10s enables precise alignment with Time-on-task-10s coding, capturing rapid cognitive fluctuations during classroom behaviors. RMSSD-60s, on the other hand, reflects broader cognitive change trends, aligning more effectively with lesson phases. Together, they provide complementary insights into both short- and long-term changes in cognitive engagement. Since RMSSD values may be affected by individual differences in the health status of participants, we performed further processing after calculating the raw RMSSD values. Specifically, we calculated the average RMSSD value for each student during the entire lesson as a baseline level of cognitive level. Subsequently, the percentage differences between RMSSD-10s and RMSSD-60s and this baseline value were calculated as RMSSD%-10s and RMSSD%-60s. By analyzing these percentage changes, we were able to integrate data from different classrooms to observe the overall cognitive change trends for all students in the classroom.

3. Results

3.1. Statistical Analysis of Code Differences and Correlations

Initially, we performed a normal distribution test on all the data. The results showed that RMSSD%-10s and RMSSD%-60s (in VLC and TLC) were normally distributed. However, the Time-on-task-10s raw coding and Time-on-task%-60s data showed a skewness distribution in both instructional settings. Therefore, in the difference analysis, we used the Kruskal-Wallis test. If there were significant differences between the categories of Time-on-task-10s (*Code 1. Private - not finished task; Code 2. Organization; Code 3. Real learning time; Code 4. Private - finished with the task*) for RMSSD%-10s, Dunn's Test was used to further compare the differences between the different codes. This approach was suitable for distinguishing between group differences in Time-on-task 10s (categorical variable). For the correlation analysis between RMSSD%-60s and Time-on-task%-60s, we used Spearman's correlation coefficient. This method is suitable for detecting the relationship between Time-on-task%-60s (continuous variables) and RMSSD.

Table. 2. *Kruskal-Wallis Test Results for Time-on-Task-10s Codes and RMSSD%-10s in VLC and TLC*

Kruskal-Wallis test	*H*	*P*
VLC	36.030	<0.001
TLC	44.426	<0.001

H= Kruskal-Wallis test value.

The Kruskal-Wallis test results (Table 2) showed that there were differences between the RMSSD%-10s in the Time-on-task-10s codes of VLC and TLC. Accordingly, Dunn's test was conducted to further analyze these differences.

Table. 3. *Sample Characteristics Table*

	VLC			TLC		
	N	$M_{(RMSSD\%\text{-}10s)}$	$SD_{(RMSSD\%\text{-}10s)}$	*N*	$M_{(RMSSD\%\text{-}10s)}$	$SD_{(RMSSD\%\text{-}10s)}$
Code 1	1106	4.591	44.265	1965	3.221	42.064
Code 2	1745	0.112	39.291	1808	-2.251	45.434
Code 3	11183	-1.237	37.967	11807	-0.600	41.538
Code 4	529	6.938	48.230	111	23.026	55.805

Table 3 shows the sample size (N), mean (M), and standard deviation (SD) of RMSSD%-10s for different Time-on-task codes. The results showed that Code 3 (real learning time) occupied the main proportion in both VLC and TCL. This indicates that students devote most of their classroom time to learning tasks. In both classrooms, Code 3's RMSSD%-10s was negative, indicating a lower heart rate variability than normal, possibly due to increased cognitive resource investment in task completion. However, RMSSD%-10s in the VLC was lower than that in the TCL, possibly indicating that students' cognitive resource investment was higher when they were on task in the VLC.

The proportion of Code 1 (Private-not finished task) was low in both classrooms, but its RMSSD%-10s value showed that students were relatively relaxed at this stage of cognition. The sample size of Code 4 (Private-finished task) was very small in both classrooms, but its RMSSD%-10s mean value was significantly higher than that of other stages. This indicates that students experienced noticeable relaxation after completing tasks. It needs to be noted that RMSSD%-10s under Code 2 (Organization) was almost the same as the baseline of students' HRV in VLC, while it was slightly lower in TCL. This may be related to the frequency of classroom organization activities. Overall, the standard deviation of TCL in Code 2, Code 3, and Code 4 were higher than those of VLC. This indicates that the cognitive state of students in TCL fluctuates more significantly, which may be affected by the classroom organization structure or teaching strategies.

Table. 4. *Dunn's Test Results for Time-on-Task-10s Codes and RMSSD%-10s in VLC and TLC*

	VLC		TLC	
Comparison	*p*	*Cohen's d*	*p*	*Cohen's d*
Code 1- Code 2	0.101	0.108	<0.001	0.125
Code 1- Code 3	< 0.001	0.151	<0.001	0.092
Code 1- Code 4	0.682	-0.051	0.002	-0.462

Code 2 - Code 3	1.00	0.035	0.124	-0.039
Code 2 - Code 4	0.002	-0.164	<0.001	-0.548
Code 3 - Code 4	< 0.001	-0.212	<0.001	-0.567

As can be seen from Table 4, there are significant differences in RMSSD%-10s under different Time-on-task codes, both in VLC and TLC. Among them, the difference between Code 3 and Code 4 was the most significant, and the effect size in TLC was much larger. Code 1 and Code 3 differed in both classrooms. Although they both showed smaller effect sizes, it was slightly higher in the VLC. In addition, the differences in RMSSD%-10s among the various codes in VLC were relatively small, indicating that the cognitive fluctuations caused by VLC were relatively gentle. However, in TLC, students' cognitive load changes were more drastic. The paired comparison results listed in Table 4 further support this finding.

To further explore the relationship between student behavior and cognitive load, we next analyzed the correlation between Time-on-task%-60s and RMSSD%-60s.

Table. 5. *Spearman Correlation Analysis Results*

	N	*Spearman Correlation (ρ)*	*p*
VLC	229	-0.1621	0.014
TLC	244	-0.2184	＜0.001

The results in Table 5 show that there is a significant negative correlation between Time-on-task%-60s and RMSSD%-60s in both classroom environments, but the strength of the correlation is different. In VLC, the correlation coefficient was weaker (ρ = -0.1621, p = 0.014), which may indicate that the guiding nature of the video stabilizes the learning state of students to some level. The cognitive level of students when using video learning fluctuates less, which reduces the dynamic link between RMSSD% and On-Task. This may have further smoothed out the cognitive fluctuations of students.

In contrast, the correlation coefficient in TLC is stronger (ρ = -0.2184, p < 0.001), indicating that students' RMSSD%-60s fluctuations are more closely related to changes in On-Task. Due to the lack of video guidance, students may need to make more autonomous efforts to complete the task. This may lead to uneven fluctuations in cognitive resource investment. In addition, teacher intervention in TLC is more frequent. This may also increase students' switching between focus and distraction, which in turn leads to greater RMSSD%-60s fluctuations.

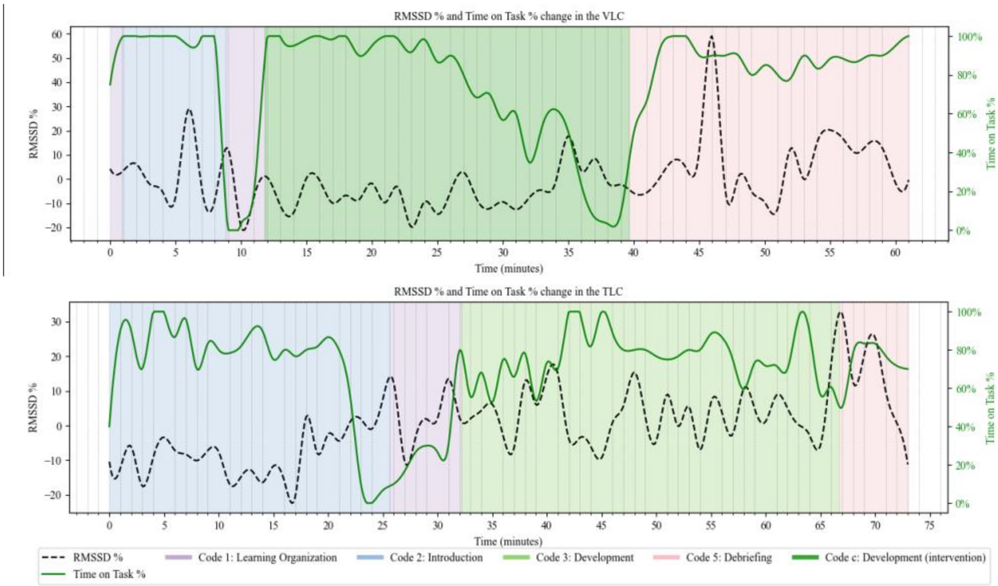
In general, there is a negative correlation between Time-on-task%-60s and RMSSD%-60s. This indicates that when students have a higher on-task percentage, it corresponds to a lower RMSSD%. The cognitive resource in VLC fluctuates more gently, while it is more drastic in TLC, reflecting the different effects of the two teaching methods on students' learning states. In the next section, we will visually compare and analyze each group of classroom data.

3.2. Visualization of Data in each group

In this section, we compare the Time-on-task%-60s and RMSSD%-60s of the entire class visually. To facilitate comparison, the results of the four groups are displayed using the same method. Figures 1, 2, 3, and 4 show the complete Time-on-task%-60s and RMSSD%-60s smoothed curves for each group in the two classrooms, with the background of each lesson phase marked in a different color.

Group A

Fig. 1. *Overall Dynamics of Time on Task%-60s and RMSSD%-60s in Group A (N=10)*

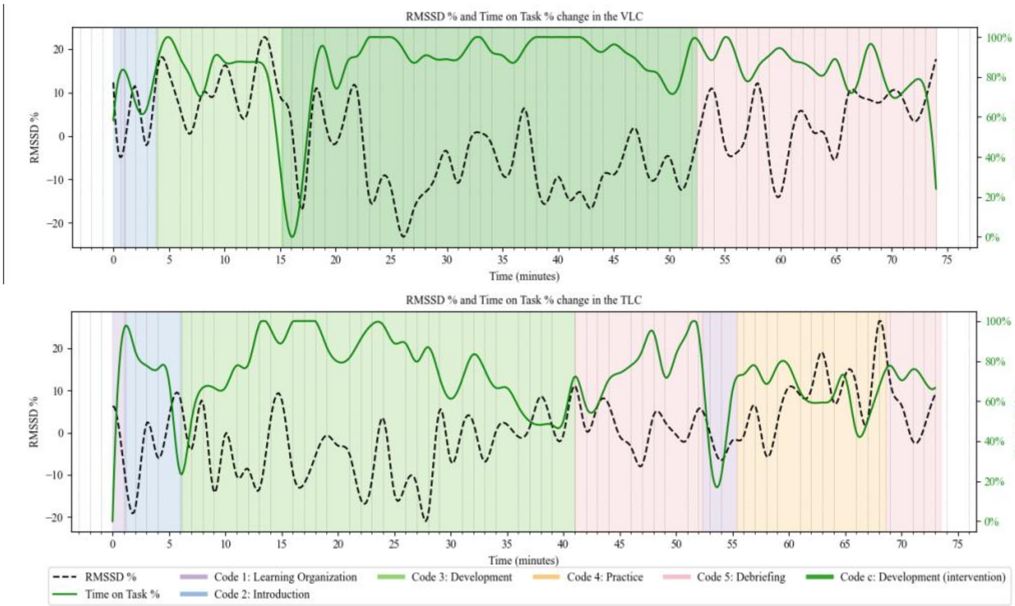


As shown in Figure 1, in VLC, the teacher first started the class with a short organization and introduction phase. At this point, students maintained a high on-task percentage. However, during the short subsequent organization time, students' on-task% decreased. This may be because the teacher was explaining the requirements of the next learning task. As the class entered the development phase, students' on-task% quickly returned to a peak, indicating that students were focused on watching the video. When the students are at the end of the development phase, the on task% begins to decrease, and the RMSSD% increases slightly. In the final debriefing phase, the students on task% returns to a higher level. However, in this phase we can see that the students' RMSSD% begins to fluctuate, especially at the 46th minute. Although it shows a high on-task%, the students are more relaxed. In the classroom recording, we observed that the teacher asked different students questions at this stage.

In TLC, the teacher started the class directly with the introduction phase. However, the students' on task% fluctuated significantly during this lesson. During this phase, the RMSSD% of the students was at a relatively low level. As time goes by, the RMSSD% of the students gradually increases, while the on task% gradually decreases. In the following organization phase, the teacher divided the students into different learning groups. The on task% of the students gradually increases, but so does the RMSDD%. In the subsequent development phase, the RMSSD% of the students remained at a fluctuating state. This may be due to the fact that students were more disturbed during group activities.

Group B

Fig. 2. *Overall Dynamics of Time on Task%-60s and RMSSD%-60s in Group B (N=9)*

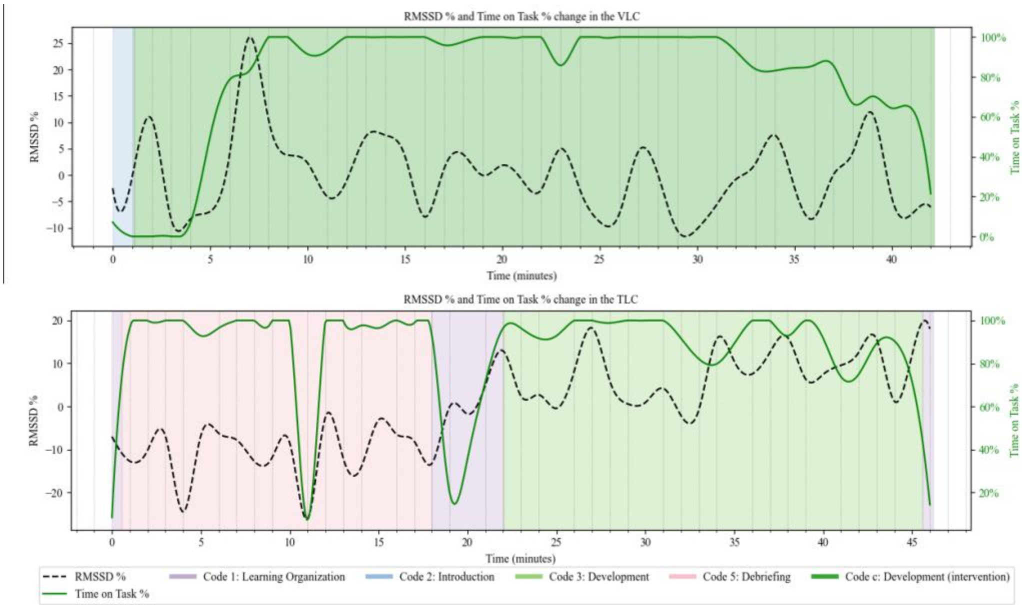


As shown in Figure 2, in the VLC classroom, the teacher began the class with a short organization and introduction phase. During these two phases, the students' on task% fluctuated relatively little and remained at a high level. When entering the development phase without video, there was a small change in the students' time-on-task, but the RMSSD% was stable. However, when entering the development phase using the video, there was a significant drop in on task%. By looking at the classroom recording, we found the students were organizing their headphones and tablets in preparation for watching the video. We noticed that after a few minutes, the students' on task% returned to a higher level and continued. The RMSSD% began to decline, which may be related to the video guide, which helps students focus more on the task.

In TLC classrooms, teachers also began the class with an organization and introduction phase. The students' on task% fluctuated from low to high and then back down again during this phase. RMSSD% was relatively low during this phase, indicating that students were investing high cognitive resources in paying attention. After entering the development phase, on task% generally increased, but fluctuated greatly, while RMSSD% showed a continuous fluctuating change. By observing the classroom recording, we found that students were independently reading the teaching materials during this phase. However, we found that some students frequently communicated with their peers, which may have been a distraction to the learning process.

Group C

Fig. 3. *Overall Dynamics of Task%-60s and RMSSD%-60s in Group C (N=14)*

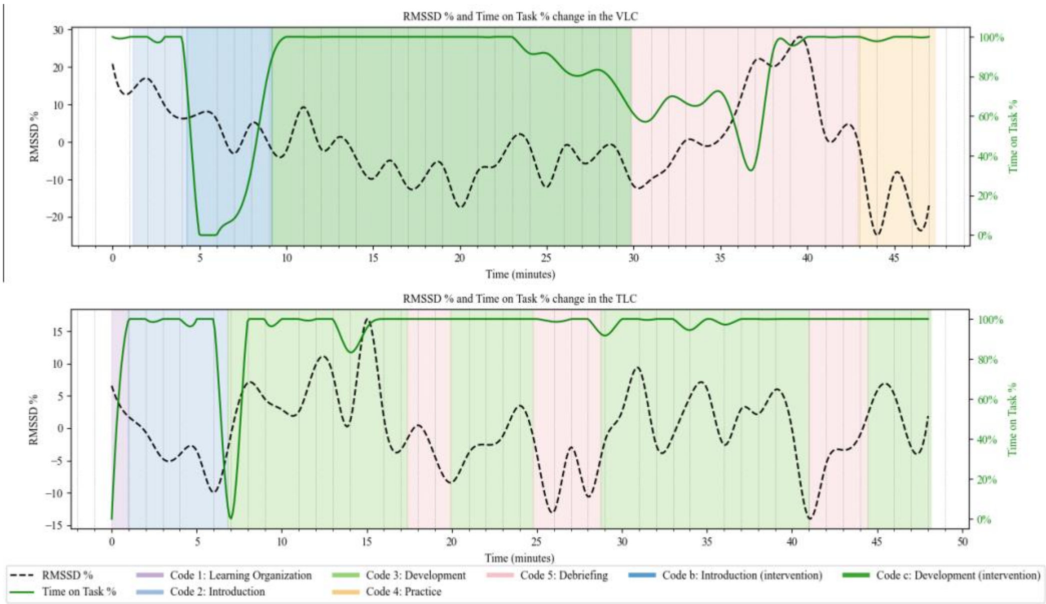


As can be seen in Figure 3, in the VLC classroom, the teacher only used 1 minute to complete the introduction phase. The students then spent the rest of the lesson working independently using the video tutorials. We observed that the students' on task% was very low during the first three minutes of the development phase, and gradually increased from the fourth minute. The RMSSD% initially experienced a brief decline before rapidly increasing. It then began to decline around the seventh minute. In the classroom recording, we observed that this was the stage where the students were preparing to watch the video. Once the students were actually engaged in watching the video, both the on task% and RMSSD% remained at a relatively stable level.

In the TLC, we can see that the teacher began the debriefing phase after organizing the class. During this phase, students took turns to present their learning outcomes and interacted with their peers and the teacher. The RMSSD% of students in this phase was at a relatively low level, indicating that students were actively participating in the class. However, as time progressed to the development phase, although the on task% remained at a relatively high level, the RMSSD% of students increased. This may indicate that students had already used up a lot of cognitive resources in the first half of the class and were already tired when it came to learn new content.

Group D

Fig. 4. *Overall Dynamics of Time on Task%-60s and RMSSD%-60s in Group D (N=12)*



As shown in Figure 4, in the VLC, the teacher began the class with the introduction phase. When the lesson entered the development phase, the students' on task% reached a peak. RMSSD showed a slow decline during this phase, which may indicate that the students were using more cognitive resources. The fluctuation remained in a relatively stable state. As the learning task was completed, we observed that the students' on task% slowly decreased, while RMSSD slowly increased. During the debriefing phase, the students were in a very relaxed state.

In TLC, the teacher alternated between the development and debriefing phases after the organization and introduction phases. We observed that, in contrast to the other classrooms, the students are in a very high on task% state at all times. In the classroom recording, we observed that the teacher alternated between explanation and interaction with the students. This shows that this teaching strategy does indeed fully engage the students. However, it is interesting to note that the students' RMSSD% always decreased during the debriefing phase. This may indicate that the students mobilized more cognitive resources during interaction with the teacher.

4. Discussion

Based on the above research results, we can answer the research questions one by one:

RQ1: Is there a difference in HRV changes for different types of tasks in the VLC and the TLC?

The results of the statistical analysis showed that there were significant differences in RMSSD%-10s under different Time-on-task codes, whether in VLC or TLC. The effect sizes between different codes in TLC was large, reflecting the fact that students in traditional classrooms may experience more frequent changes in cognitive. In addition, when the time-on-task was Code 3 (real learning time), a lower RMSSD% was shown in VLC. Our results align with previous research findings on cognitive load and HRV (Haapalainen et al., 2010; Mizuno et al., 2011; Minkley et al., 2018; Mukherjee et al., 2011). When students experience higher cognitive load, HRV decreases. This indicates that the video content had an effect on increasing students' cognitive load during the learning task. It can also be seen in the visualization results that the HRV curve tended to be lower than the classroom average during the developmental phase of VLC. Students may have had a higher cognitive load during this phase.

RQ2: Is there a correlation between HRV changes and Time-on-task changes in VLC and TLC?

The Spearman correlation analysis shows a negative correlation between Time-on-task%-60s and RMSSD%-60s. This correlation is significant in both classroom settings, but the strength differs. The correlation in the VLC classroom is weaker (ρ = -0.1621). In contrast, changes in Time-on-task% in the TLC classroom had a greater impact on fluctuations in RMSSD%, with a stronger negative correlation (ρ = -0.2184). This may be because the interaction and task design in traditional classrooms place higher demands on students' self-regulatory abilities. Especially when working in groups, students need to switch attention and allocate cognitive resources frequently, resulting in more significant fluctuations in cognitive.

This result also shows that both time-on-task and HRV can measure students' learning states. However, the combination of the two can help to provide a more accurate depiction of student engagement. For example, by observing the visualization of the four groups, we can see that in VLC, students' HRV tended to be below average when they were using the video tutorial to complete learning tasks (development phase). This indicates that students may have experienced more cognitive load when using the video (Paas et al., 2008). During non-learning tasks, students' HRV tended to be above average. In TLC, the HRV trend was gradually upward in all classrooms except Group D. This indicates that the cognitive engagement of the students gradually decreased over time. This is particularly the case in Group C, where the teacher placed the development phase at the end, when the students' HRV levels were already above average. Even though Time-on-task showed a high level during this phase, the students were actually in a more relaxed state and did not really devote more cognitive resources. This indicates that students' cognitive resources have been depleted and they may be engaged in surface learning at this phase (Chen et al., 2018; Dolmans et al., 2016). This could not be detected by classroom observation alone. The peculiarity of Group D is that the teacher interspersed the development and debriefing phases, which resulted in a fluctuating trend in the students' HRV.

Therefore, based on the above findings, we propose the following recommendations for future instructional design. First, teachers should take advantage of multimedia technology by using videos in the instruction to guide students in allocating cognitive resources effectively. Second, teachers should avoid placing development phases at the end of lessons when designing their lessons. This is because the depletion of cognitive resources may result in students being unable to fully engage in learning towards the end of the lesson. Finally, the study found that students needed to allocate cognitive resources frequently when working on group tasks and during the debriefing phase. This suggests that teaching interventions could be used to improve students' self-regulation when no videos are used. For example, providing specific task steps could help students focus and reduce distractions.

4.1. Limitations and future work

This study also has some limitations. First, this study used RMSSD% as the main measure of cognitive engagement. Although this indicator has significant advantages in reflecting students' ANS activity, it is still a single indicator. To further enrich the understanding of the learning state, future studies can consider combining other biofeedback data for analysis.

Second, it can be observed from the visualization results that there are certain differences in task engagement and cognitive fluctuations between vocational school (Group A, Group B) and general high school (Group C, Group D) students. This may be related to the type of school or students' learning habits, but the current study was unable to further analyze the specific causes of this difference.

Finally, the study did not directly measure outcome variables such as academic performance, so it was not possible to comprehensively assess the specific impact of cognitive fluctuations on learning outcomes. Future research should explore the relationship between the dynamic changes in cognitive engagement and learning outcomes by combining learning outcome data.

4.2. Conclusions

Our study, based on Multimodal Learning Analytics, investigated the dynamic relationship between Time-on-task and HRV in two instructional settings: VLC and TLC. By integrating behavioral and biofeedback data, this study revealed the characteristics of students' learning states in different classroom settings. The results showed that time-on-task was negatively correlated with HRV, and that combining the two provided a more comprehensive understanding of students' learning states. Students' cognitive states were relatively stable in the VLC classroom, especially during the development phase. However, students' cognitive level fluctuated more in the TLC classroom. This indicates that video tutorials play an important role in mobilizing students' cognitive resources and reducing fluctuations in cognitive load. The research results provide empirical support for the optimization of instructional design. In addition, our study verifies the potential of multimodal learning analytics in real classroom settings.

Author Contributions

All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Enqi Fan and Jens Siemon. The first draft of the manuscript was written by Enqi Fan. Matt Bower, Jens Siemon revised it critically for important intellectual content. All authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Finding

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

References

Anderson, L. W. (1995). International encyclopedia of teaching and teacher education. Elsevier Science Inc., 660 White Plains Road, Tarrytown, NY 10591-5153.
Antonenko, P., Paas, F., Grabner, R., & van Gog, T. (2010). Using Electroencephalography to Measure Cognitive Load. Educational Psychology Review, 22(4), 425–438. https://doi.org/10.1007/s10648-010-9130-y
Ayres, P. (2006). Using subjective measures to detect variations of intrinsic cognitive load within problems. Learning and Instruction, 16(5), 389–400. https://doi.org/10.1016/j.learninstruc.2006.09.001
Ayres, P., Lee, J. Y., Paas, F., & Van Merrienboer, J. J. (2021). The validity of physiological measures to identify differences in intrinsic cognitive load. Frontiers in psychology, 12, 702538.
Bernardi, L., Wdowczyk-Szulc, J., Valenti, C., Castoldi, S., Passino, C., Spadacini, G., & Sleight, P. (2000). Effects of controlled breathing, mental activity and mental stress with or without verbalization on heart rate variability. Journal of the American College of Cardiology, 35(6), 1462–1469. https://doi.org/10.1016/s0735-1097(00)00595-7
Borko, H., Jacobs, J., Eiteljorg, E., & Pittman, M. E. (2008). Video as a tool for fostering productive discussions in mathematics professional development. Teaching and Teacher Education, 24, 417–436. https://doi.org/10.1016/j.tate.2006.11.012
Buijs, M., & Admiraal, W. (2013). Homework assignments to enhance student engagement in secondary education. European journal of psychology of education, 28, 767-779. https://doi.org/10.1007/s10212-012-0139-0
Carroll, J. B. (1963). A model of school learning. Teachers college record, 64(8), 1-9.
Carvalho, P.F., Mclaughlin, E., & Koedinger, K. (2017). Is there an explicit learning bias? Students beliefs, behaviors and learning outcomes. Cognitive Science.
Chen, O., Castro-Alonso, J. C., Paas, F., & Sweller, J. (2018). Extending cognitive load theory to incorporate working memory resource depletion: Evidence from the spacing effect. Educational Psychology Review, 30, 483-501. https://doi.org/10.1007/s10648-017-9426-2
Corno, L., & Mandinach, E. B. (1983). The role of cognitive engagement in classroom learning and motivation. Educational Psychologist, 18(2), 88–108. https://doi.org/10.1080/00461528309529266
Cowling, M., & Birt, J. (2020). Mixed reality multimodal learning analytics. Encyclopedia of Educational Innovation, 1–7. https://doi.org/10.1007/978-981-13-2262-4_149-1
Cukurova, M., Giannakos, M., & Martinez-Maldonado, R. (2020). The promise and challenges of multimodal learning analytics. British Journal of Educational Technology, 51(5), 1441-1449. https://doi.org/10.1111/bjet.13015
Dignath, C., and Veenman, M. V. J. (2020). The role of direct strategy instruction and indirect activation of self-regulated learning–evidence from classroom observation studies. Educ. Psychol. Rev. 33, 489–533. https://doi.org/10.1007/s10648-020-09534-0
Dolmans, D. H., Loyens, S. M., Marcq, H., & Gijbels, D. (2016). Deep and surface learning in problem-based learning: a review of the literature. Advances in health sciences education, 21, 1087-1112. https://doi.org/10.1007/s10459-015-9645-6
Egger, M., Ley, M., & Hanke, S. (2019). Emotion recognition from Physiological Signal Analysis: A Review. Electronic Notes in Theoretical Computer Science, 343, 35–55. https://doi.org/10.1016/j.entcs.2019.04.009
Elkin, R. L., Beaubien, J. M., Damaghi, N., Chang, T. P., & Kessler, D. O. (2024). Dynamic Cognitive Load Assessment in Virtual Reality. Simulation & Gaming, 55(4), 755-775. https://doi.org/10.1177/10468781241248821
Fan, E. & Siemon, J. (2024). Classroom Observation Coding Manual System - Lesson Phases, Social Forms, and Time-on-Task [Ratingmanual: Version 1.0]. Erstanwendung 2015. Frankfurt am Main: Forschungsdatenzentrum Bildung am DIPF. http://dx.doi.org/10.7477/1152:415
Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. Behavior Research Methods, 41(4), 1149–1160. https://doi.org/10.3758/brm.41.4.1149
Fischer, H.E., Neumann, K. (2012). Video Analysis As A Tool For Understanding Science Instruction. In: Jorde, D., Dillon, J. (eds) Science Education Research and Practice in Europe. Cultural Perpectives in Science Education, vol 5. SensePublishers, Rotterdam. https://doi.org/10.1007/978-94-6091-900-8_6
Fisher, C. W. (1978). Teaching Behaviors, Academic Learning Time and Student Achievement: Final Report of Phase III-B, Beginning Teachers Evaluation Study. Technical Report V-1
Forte, G., Favieri, F., & Casagrande, M. (2019). Heart rate variability and cognitive function: a systematic review. Frontiers in neuroscience, 13, 710. https://doi.org/10.3389/fnins.2019.00710
Fredrick, W. C., Walberg, H. J., & Rasher, S. P. (1979). Time, Teacher Comments, and Achievement in Urban High Schools. The Journal of Educational Research, 73(2), 63–65. https://doi.org/10.1080/00220671.1979.10885209
Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School Engagement: Potential of the Concept, State of the Evidence. Review of Educational Research, 74(1), 59-109. https://doi.org/10.3102/00346543074001059
Gettinger, M., & Seibert, J. K. (2002). Best Practices in Increasing Academic Learning Time. In A. Thomas & J. Grimes (Eds.), Best practices in school psychology IV (pp. 773–787). National Association of School Psychologists. https://psycnet.apa.org/record/2006-03715-049
Ghergulescu, I., & Muntean, C. H. (2016). ToTCompute: A novel EEG-based TimeOnTask threshold computation mechanism for engagement modelling and monitoring. International Journal of Artificial Intelligence in Education, 26(3), 821–854. https://doi.org/10.1007/s40593-016-0111-2

Giannakos, M., Spikol, D., Di Mitri, D., Sharma, K., Ochoa, X., Hammad, R. (2022). Introduction to Multimodal Learning Analytics. In: Giannakos, M., Spikol, D., Di Mitri, D., Sharma, K., Ochoa, X., Hammad, R. (eds) The Multimodal Learning Analytics Handbook. Springer, Cham. https://doi.org/10.1007/978-3-031-08076-0_1

Gilgen-Ammann, R., Schweizer, T., & Wyss, T. (2019). RR interval signal quality of a heart rate monitor and an ECG Holter at rest and during exercise. European Journal of Applied Physiology, 119(7), 1525–1532. https://doi.org/10.1007/s00421-019-04142-5

Götzl, M., Jahn, R. W., & Held, G. (2013). Bleibt alles anders!? Sozialformen, Unterrichtsphasen und echte Lernzeit im kaufmännischen Unterricht. Bwp@ Berufs- Und Wirtschaftspädagogik Online, 24, 1-22. Abgerufen unter: http://www.bwpat.de/ausgabe24/goetzl_etal_bwpat24.pdf.

Granström, M., Kikas, E., & Eisenschmidt, E. (2023). Classroom observations: How do teachers teach learning strategies?. In Frontiers in Education (Vol. 8, p. 1119519). Frontiers Media SA. https://doi.org/10.3389/feduc.2023.1119519

Grassmann, M., Vlemincx, E., von Leupoldt, A., & Van den Bergh, O. (2015). The role of respiratory measures to assess mental load in pilot selection. Ergonomics, 59(6), 745–753. https://doi.org/10.1080/00140139.2015.1090019

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

Haapalainen, E., Kim, S., Forlizzi, J. F., & Dey, A. K. (2010). Psycho-physiological measures for assessing cognitive load. Proceedings of the 12th ACM International Conference on Ubiquitous Computing. https://doi.org/10.1145/1864349.1864395

Hassib, M., Khamis, M., Friedl, S., Schneegass, S., & Alt, F. (2017, November). Brainatwork: Logging cognitive engagement and tasks in the workplace using electroencephalography. In Proceedings of the 16th international conference on mobile and ubiquitous multimedia (pp. 305-310).

Helmke, A. (2007). Aktive Lernzeit optimieren. Was wissen wir über effiziente Klassenführung? In Pädagogik (Weinheim), 59(5), 44-49.

Hesse, H. G. (1994). Lehr-Lern-Zeit und Lernerfolg aus psychologischer Sicht (pp. 143-161).

Hjortskov, N., Rissén, D., Blangsted, A. K., Fallentin, N., Lundberg, U., & Søgaard, K. (2004). The effect of mental stress on heart rate variability and blood pressure during computer work. European Journal of Applied Physiology, 92(1–2), 84–89. https://doi.org/10.1007/s00421-004-1055-z

Hong, J., Pi, Z., & Yang, J. (2016). Learning declarative and procedural knowledge via video lectures: Cognitive Load and learning effectiveness. Innovations in Education and Teaching International, 55(1), 74–81. https://doi.org/10.1080/14703297.2016.1237371

Kalyuga, S. (2011). Cognitive load theory: How many types of load does it really need?. Educational psychology review, 23, 1-19. https://doi.org/10.1007/s10648-010-9150-7

Klette, K. (2015). Introduction: Studying interaction and instructional patterns in classrooms. In K. Klette, O. K. Bergem, & A. Roe (Eds.), Teaching and learning in lower secondary schools in the era of PISA and TIMSS (pp. 1–16). Springer International Publishing.

Klette, K. (2023). Classroom observation as a means of understanding teaching quality: towards a shared language of teaching? Journal of Curriculum Studies, 55(1), 49–62. https://doi.org/10.1080/00220272.2023.2172360

Knigge, M., Duarte, J. da S., Nordstrand, V., Siemon, J., & Stolp, C. (2013). Videostudien als Mittel fachdidaktischer Erkenntnisgewinnung. bwp@ Berufs- und Wirtschaftspädagogik – online, 24. https://www.bwpat.de/ausgabe/24/knigge-etal

Levenson, R. W. (2003). Blood, sweat, and fears. Annals of the New York Academy of Sciences, 1000(1), 348–366. https://doi.org/10.1196/annals.1280.016

Li, S. (2021). Measuring cognitive engagement: An overview of measurement instruments and techniques. International Journal of Psychology and Educational Studies, 8(3), 63-76.

Lipowsky, F. (2006). Auf den Lehrer kommt es an. Empirische Evidenzen für Zusammenhänge zwischen Lehrerkompetenzen, Lehrerhandeln und dem Lernen der Schüler (pp. 47-70).

Miller, B. W. (2015). Using reading times and eye-movements to measure cognitive engagement. Educational Psychologist, 50(1), 31–42. doi:10.1080/00461520.2015.100406

Minkley, N., Kärner, T., Jojart, A., Nobbe, L., & Krell, M. (2018). Students' mental load, stress, and performance when working with symbolic or symbolic–textual molecular representations. Journal of Research in Science Teaching, 55(8), 1162–1187. https://doi.org/10.1002/tea.21446

Mizuno, K., Tanaka, M., Yamaguti, K., Kajimoto, O., Kuratsune, H., & Watanabe, Y. (2011). Mental fatigue caused by prolonged cognitive load associated with sympathetic hyperactivity. Behavioral and Brain Functions, 7(1), 17. https://doi.org/10.1186/1744-9081-7-17

Moya-Ramon, M., Mateo-March, M., Peña-González, I., Zabala, M., & Javaloyes, A. (2022). Validity and reliability of different smartphones applications to measure HRV during short and ultra-short measurements in elite athletes. Computer Methods and Programs in Biomedicine, 217, 106696. https://doi.org/10.1016/j.cmpb.2022.106696

Mukherjee S, Yadav R, Yung I, Zajdel DP, Oken BS (2011) Sensitivity to mental effort and test–retest reliability of heart rate variability measures in healthy seniors. Clin Neurophysiol 122(10):2059–2066. https://doi.org/10.1016/j.clinph.2011.02.032

Mukherjee, S., Yadav, R., Yung, I., Zajdel, D. P., & Oken, B. S. (2011). Sensitivity to mental effort and test-retest reliability of heart rate variability measures in healthy seniors. Clinical Neurophysiology. https://doi.org/10.1016/j.clinph.2011.02.032

Mullikin, D. R., Flanagan, R. P., Merkebu, J., Durning, S. J., & Soh, M. (2024). Physiologic measurements of cognitive load in clinical reasoning. Diagnosis, 11(2), 125-131. https://doi.org/10.1515/dx-2023-0143
Ouhaichi, H., Spikol, D., & Vogel, B. (2023). Research trends in multimodal learning analytics: A systematic mapping study. Computers and Education: Artificial Intelligence, 4, 100136.
Paas, F. G. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. Journal of Educational Psychology, 84(4), 429–434. https://doi.org/10.1037/0022-0663.84.4.429
Paas, F. G. W. C., Ayres, P., & Pachman, M. (2008). Assessment of cognitive load in multimedia learning. Recent Innovations in Educational Technology That Facilitate Student Learning, Information Age Publishing Inc., Charlotte, NC, 11-35.
Pham, T., Lau, Z. J., Chen, S. H., & Makowski, D. (2021). Heart rate variability in psychology: A review of HRV Indices and an analysis tutorial. Sensors, 21(12), 3998. https://doi.org/10.3390/s21123998
Rolim, L. C., de Souza, J. S., & Dib, S. A. (2013). Tests for early diagnosis of cardiovascular autonomic neuropathy: Critical analysis and relevance. Frontiers in Endocrinology, 4. https://doi.org/10.3389/fendo.2013.00173
Scholkmann, A., Siemon, J., Boom, K. D., & Knigge, M. (2017). Lernzeitnutzung im Planspielunterricht: eine Analyse des Einflusses kognitiver Fähigkeiten, Zielorientierungen und Charakteristika von Lernpartnern anhand von Videodaten. Zeitschrift für Erziehungswissenschaft, 20(4), 651-669.
Sharma, R. (2017). Computer assisted learning–A study. International Journal of Advanced Research in Education & Technology (IJARET), 4(2), 102-105.
Siemon, J., Scholkmann, A., Boom, K. D., & Hamburg, U. (2015). Time on Task in Collaborative Learning. Influence of Learning Goal Orientation and Group Composition.
Solhjoo, S., Haigney, M. C., McBee, E., van Merrienboer, J. J., Schuwirth, L., Artino, A. R., Battista, A., Ratcliffe, T. A., Lee, H. D., & Durning, S. J. (2019). Heart rate and heart rate variability correlate with clinical reasoning performance and self-reported measures of cognitive load. Scientific Reports, 9(1). https://doi.org/10.1038/s41598-019-50280-3
Song, H. S., & Lehrer, P. M. (2003). The effects of specific respiratory rates on heart rate and heart rate variability. Applied psychophysiology and biofeedback, 28, 13-23. https://doi.org/10.1023/A:1022312815649
Spanjers, D. M., Burns, M. K., & Wagner, A. R. (2008). Systematic direct observation of time on task as a measure of student engagement. Assessment for Effective Intervention, 33(2), 120-126.
Speer, K. E., Semple, S., Naumovski, N., & McKune, A. J. (2020). Measuring heart rate variability using commercially available devices in healthy children: A validity and Reliability Study. European Journal of Investigation in Health, Psychology and Education, 10(1), 390–404. https://doi.org/10.3390/ejihpe10010029
Stein, P. K., Bosner, M. S., Kleiger, R. E., & Conger, B. M. (1994). Heart rate variability: A measure of cardiac autonomic tone. American Heart Journal, 127(5), 1376–1381. https://doi.org/10.1016/0002-8703(94)90059-0
Sweller, J. (1988). Cognitive Load During Problem Solving: Effects on Learning. Cognitive Science, 12(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4
Sweller, J., van Merrienboer, J. J., & Paas, F. G. (1998). Cognitive architecture and instructional design. Educational Psychology Review, 10(3), 251–296. https://doi.org/10.1023/a:1022193728205
Thong, T., Li, K., McNames, J., Aboy, M., & Goldstein, B. (2003). Accuracy of ultra-short heart rate variability measures. Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEEE Cat. No.03CH37439). https://doi.org/10.1109/iembs.2003.1280405
Vitolo, E., Giromini, L., Viglione, D. J., Cauda, F., & Zennaro, A. (2021). Complexity and cognitive engagement in the Rorschach task: An fMRI study. Journal of Personality Assessment, 103(5), 634-644.
Xiong, R., Kong, F., Yang, X., Liu, G., & Wen, W. (2020). Pattern recognition of cognitive load using EEG and ECG signals. Sensors, 20(18), 5122. https://doi.org/10.3390/s20185122
Ziemssen, T., & Siepmann, T. (2019). The investigation of the cardiovascular and sudomotor autonomic nervous system—a review. Frontiers in neurology, 10, 53. https://doi.org/10.3389/fneur.2019.00053

Appendix A.

A.1. Classroom Observation Coding Manual System- Lesson Phases, Social Forms, and Time-on-Task

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

FAKULTÄT
FÜR ERZIEHUNGSWISSENSCHAFT

Classroom Observation Coding Manual System

- Lesson Phases, Social Forms, and Time-on-Task

Classroom observation based on video data

Authors

Enqi Fan
Faculty of Education University of Hamburg
Hamburg, Germany

Jens Siemon
Faculty of Education University of Hamburg
Hamburg, Germany

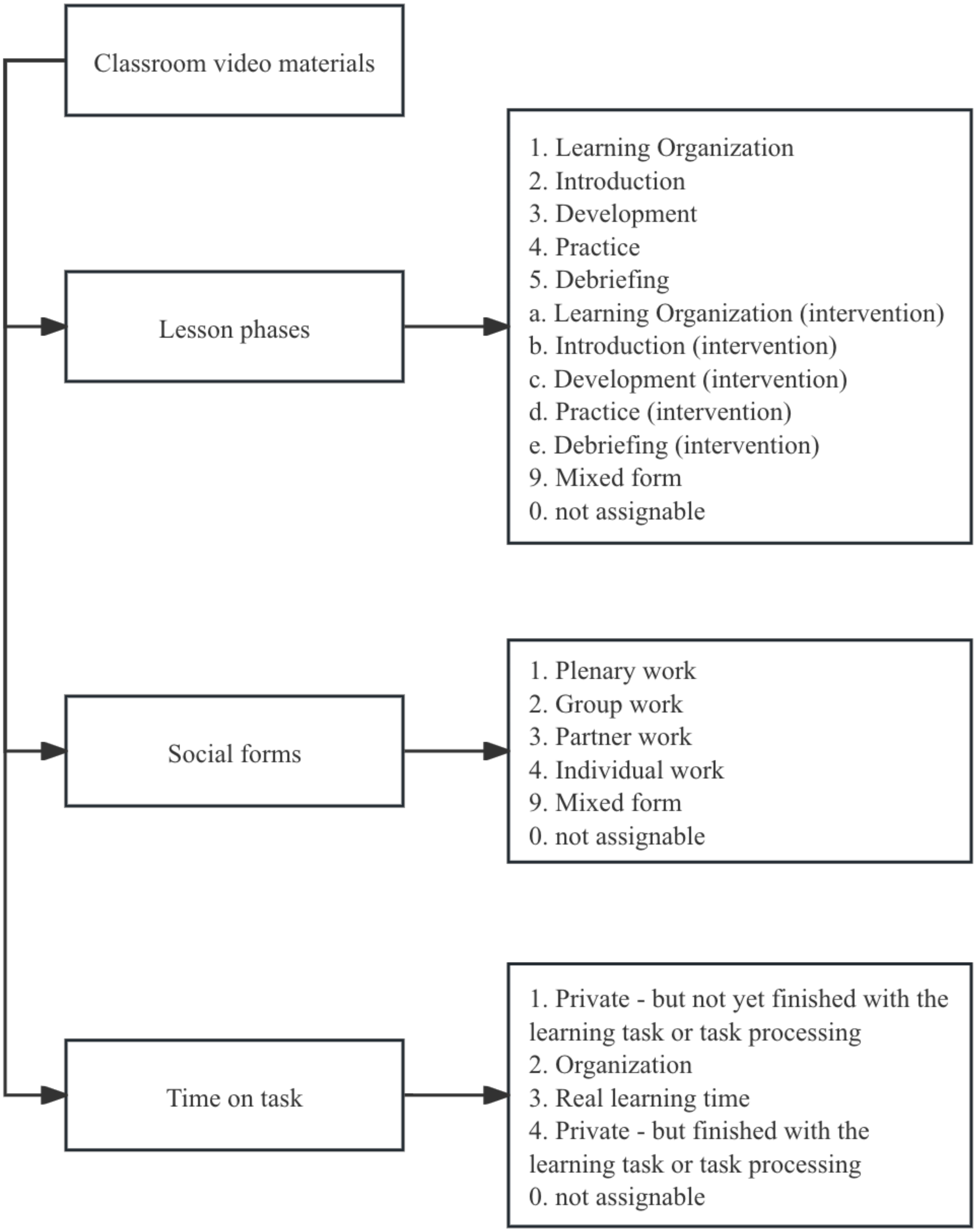
May 21, 2024

Introduction

In classroom observation research, video analysis has been a powerful tool for enhancing the quality of instruction (Klette, 2009; Fischer & Neumann, 2012; Klette, 2023). Observing and analyzing classroom videos can allow researchers to explore instructional activities and interactions in greater detail (Stigler, 1997; Seidel, 2005; Derry et al., 2010). However, many classroom observations are based on non-standardized, informal, unvalidated tools and these subjective preferences can make the analysis unreliable (Klette & Blikstad-Balas, 2018; Bostic, 2021). Therefore, in order to improve the accuracy and reliability of classroom observation, developing a coding manual is necessary (Schoenfeld, A. H. 2013; Klette & Blikstad-Balas, 2018).

Several studies have been conducted to analyze classrooms through coding (Seidel, 2005; Pianta 2008; Siemon, 2018). Based on the previous studies (Seidel, 2005; Beerenwinkel & Börlin, 2014; Siemon, 2015) we developed and revised the classroom observation coding manual. This has resulted in a coding manual system that enables comprehensive observation of the classroom from three perspectives: lesson phases, classroom social forms, and time-on-task. This coding system contains three coding manuals. Here we first provide an overview of the relationship between the three coding manuals, then describe the inter-rater agreement (IRA) and inter-rater reliability (IRR) of coding manuals and the training of coders. Finally, we provide a detailed description of the three coding manuals, each of which contains a detailed set of behavioral standards and categorical event coding rules.

Overview of Coding Manual Relationships



Inter-Rater Agreement and Inter-Rater Reliability of Coding Manuals

To ensure the validity of our coding system, we conducted Inter-Rater Agreement (IRA) and Inter-Rater Reliability (IRR) assessments of three coding manuals. (Chaturvedi & Shweta, 2015).
We use Cohen's kappa to measure inter-rater agreement (Cohen, 1960; Grouven, 2007). Two coders independently coded the classroom videos and calculated Cohen's kappa values. A high kappa value indicated that the two coders had a high level of agreement in using the coding manual (Table 1).

Table 1.

Interpretation of kappa coefficient

kappa coefficient	Interpretation
k<0	No agreement
0<k<0,2	Slight agreement
0,21<k<0,4	Fair agreement
0,41<k<0,6	Moderate agreement
0,61<k<0,8	Substantial agreement
0,81<k<1	Almost perfect agreement

Meanwhile, to evaluate the reliability of the coding manuals, we coded multiple classrooms and calculated the kappa value for each classroom. In Table 2, we provide a range of kappa values for the three coding manuals. This range reflects the reliability of the coding manuals in different classrooms and different combinations of coders.

Table 2.

Inter-rater reliability

Coding manual	Cohens Kappa
lesson phases	0,71-1,00
social forms	0,86-1,00
time-on-task	0,78-0,90

Training of coders

In order to ensure the quality of coding results, we recommend training coders before they code research data. First, coders need to read the coding manual in detail. Group discussions can also be held to ensure that all coders understand the coding rules and categorization standards. Secondly, coders need to understand the coding software and practice coding with sample videos. We use the video analysis software Mangold Interact (Mangold International GmbH, Arnstorf, Germany) to code the classroom videos. In this stage, two coders work as a group to independently code the same classroom video. Once the coding is complete, coders will compare each other's coding results, discuss inconsistencies, and share their understanding of the coding rules. This practice process can be repeated several times until the two coders reach a standard of agreement across multiple independent coding sessions. After this, coders can begin working on independent coding.

Coding Manual - Lesson Phases

Coding guidelines

Definition of lesson phases

Lesson phases are the different stages in the teaching and learning process (Beerenwinkel & Börlin, 2014).

Instructions of use

Observation and coding should start with a clearly recognizable action, e.g. when the teacher starts to introduce the topic or to create groups for students. For example, from timestamp 00:01:21 the teacher starts to introduce the topic of the lesson, until timestamp 00:03:35 when the teacher ends the phase. In this situation, we can code this phase as the introduction. The next phase should directly start at timestamp 00:03:36. There should be no gaps or intervals period between each phase, this is because intervals are students or teachers wrapping up the content or activity of the previous phase. Following this, the teacher will start a new phase through clear instructions and actions. Coding ends when all instructional activities are completed.

Notes on Saving Coding Files

We recommend a unique naming for each coding file, which should include the name of the recording, the observation number, and the name of the coding manual, e.g., VidXXX_01_lesson_phases.

Lesson phases events and demarcation

In the Lesson Phases Coding Manual, we defined 12 different categories.
In these categories, in addition to the base codes, we developed codes for classroom intervention events (from codes *a* to *e*). Researchers can conduct a comparative analysis between regular classrooms and intervention classrooms through the use of video tutorials, virtual reality, games, or other activities to intervene in instruction.

1. Learning Organization
2. Introduction
3. Development
4. Practice
5. Debriefing
a. Learning Organization (intervention)
b. Introduction (intervention)
c. Development (intervention)
d. Practice (intervention)
e. Debriefing (intervention)
9. Mixed form
0. not assignable

Category system

Code	Category	Definition	Examples	Delimitation
1	**Learning Organization**	The teacher organizes activities which are not related to the content of the lesson topic.	• Welcoming students • Explaining the lesson timetable • Establishing spatial structure, e.g., seat assignment • Creating social structures, e.g. grouping students • Plan next steps	-Phase contains no knowledge content. -Does not include references to content that has already happened in the classroom, as this is part of the introduction.
2	**Introduction**	The teacher starts explaining course topic but has not yet communicated the new content of the lesson.	• Asking what the topic means • Categorizing the topic • Continuing from previous lessons • Utilizing students' general knowledge or experiential knowledge • Sensitizing students to the new topic • Activation through media (film, introductory story)	-The content to be learned in this lesson has not yet been covered. -It does not relate to learning organizations and must be relevant to the lesson content.

			• Repeating and reviewing previously taught knowledge • Asking students how somebody managed this situation • Asking what if ...	
3	**Development**	Let students learn and familiarize themselves with new content.	• Reading an instructional text • The teacher explains the new lesson (may ask questions) • Work together to create a blackboard	-The goal is to acquire new knowledge, not to practice what students already know.
4	**Practice**	Students apply what they have already learned.	• Processing worksheets • Preparing a presentation • Completing learning materials	-Students have a high level of self-directed activity in individual, partner, or group work. No plenary control by the teacher.
5	**Debriefing**	Communication between the teacher and students regarding the results of the lesson.	• Students report the results of the practice (e.g., Blackboard) • Summarize or check learning outcomes (e.g., presentation or poster).	
a	**Learning Organization (intervention)**	The teacher organizes activities which are not related to the content of the lesson.	• The teacher organizes the classroom through interventions (e.g., grouping the classroom through games)	
b	**Introduction (intervention)**	The teacher starts to talk about the course topic but has not yet communicated the new content of the lesson.	• The teacher introduces the topic to be covered in the lesson through a video tutorial or game	-Stimulate students' curiosity through a lead-in. Get students into a state of learning without involving new content.

c	**Development (intervention)**	Let students learn and familiarize with new content through interventions.	• Introducing new knowledge through video tutorials • Students have the opportunity to view video tutorials individually and complete learning tasks	-Explore whether interventions can make new content easier for students to understand.
d	**Practice (intervention)**	Students apply previous learning through interventions.	• Students apply previous learning and strengthen their understanding actively through targeted practice, supported by video tutorials/games/VR	-Explore whether interventions can enhance this practice phase for students by helping to strengthen the knowledge and skills students have already learned.
e	**Debriefing (intervention)**	Checking or comparing student solutions through intervention.	• Students compare their results and solutions with those in the video tutorials in order to check that they have understood and applied the learning correctly.	-Interventions should be used as references to give students the opportunity to examine and evaluate their own solutions.
9	**Mixed form**	This categorization is unlikely to happen in the classroom		
0	**not assignable**	This is also unlikely to happen.		

Coding Manual - Social Forms

Coding guidelines

Definition of social forms

Social forms are the specific organization of classroom interactions, these describe the social relationships and patterns of communication among classroom participants. This includes interactions between students, between students and their teacher, and group combinations. Depending on how the classroom is organized, social forms can be individual work, partner work, group work, or plenary work (Euler & Hahn, 2014).

Instructions of use

Observation and coding should start with a clearly recognizable action, e.g. when the teacher starts to introduce the topic or to create groups for students. For example, from timestamp 00:00:35, the teacher begins to introduce the lesson topic to all students and informs them of the next phase of group work. By timestamp 00:05:27, the plenary work ends, this is because the first group or student starts group work at timestamp 00:05:28. There should be no gaps or intervals between each phase. The period before timestamp 00:05:28, even if it is a transition time, should be included in the plenary work. This is because the transition time is actually students or teachers wrapping up the content or activity of the previous phase. Coding ends when all instructional activities are completed.

Notes on Saving Coding Files

We recommend a unique naming for each coding file, which should include the name of the recording, the observation number, and the name of the coding manual, e.g., VidXXX_01_social_forms.

Social forms events and demarcation

1. Plenary work
2. Group work
3. Partner work
4. Individual work
9. Mixed form
0. not assignable

Category system

Code	Category	Definition	Examples	Delimitation
1	**Plenary work**	All students follow the teacher's instructions.	• Lecture by the teacher • The teacher holds a discussion with the students on a particular topic. • The class brainstorms together on a topic.	
2	**Group work**	In a group of > two students, the students control the (learning) activities.	• Students sit at group tables and solve a task.	-Number of students greater than two
3	**Partner work**	In a group of two students, the students control the (learning) activities.	• Two students sit next to each other and work in pairs on a task.	-Only two students are allowed
4	**Individual work**	Each student is in control of their own (learning) activities.	• Students work alone on a task.	
9	**Mixed form**	This categorization is unlikely to happen in the classroom		-Group work that also involves partner or individual work is categorized as group work. -Partner work that also involves individual work is

				categorized as partner work. -About Plenary work: in principle, all students are under the control of the teacher's actions, e.g. listening to the teacher's lectures. If at this phase individual students have other tasks, this is group work. In this case, it means that the teacher divides the class into groups and working in different ways.
0	**not assignable**	This is also unlikely to happen.		

Coding Manual - Time-on-Task

Coding guidelines

Definition of time-on-task

Time-on-task is the amount of time that a student is engaged in active learning for the completion of a learning task (Hesse, 1994).

Instructions of use

Observation and coding should start with a clearly recognizable action, e.g. when the teacher starts to introduce the topic or to create groups for students. Coding ends when all instructional activities are completed.
Individual students get rated in **10-second intervals**, which means they are continuously rated. The timeline is divided into 10-second intervals and for each interval a new marker is defined. The first marker goes from timestamp 00:00:01 to timestamp 00:00:10, the second marker goes from timestamp 00:00:11 to timestamp 00:00:20, and so on. Each 10-second interval is evaluated to determine which of the major observable activities in the table below characterize this period.

Notes on Saving Coding Files

We recommend a unique naming for each coding file, which should include the name of the recording, the observation number, and the name of the coding manual, e.g. VidXXX_01_time_on_task.

Time-on-task themes and typical cases

1. Private - but not yet finished with the learning task or task processing
2. Organization
3. Real learning time
4. Private - but finished with the learning task or task processing
0. not assignable

Table 1: Coding Scheme Time-on-Task (TT)

Theme	Typical cases
Private - **but not yet finished** with the learning task or task processing [1]	Playing with cell phones or other devices while interrupting learning tasks
	Talking about personal matters with peers while interrupting learning tasks
	Talking about personal matters with the teacher while interrupting learning tasks
	Talking about personal matters with other group members while interrupting learning tasks
	Does not pay attention to the lesson and keeps to himself/herself while interrupting the learning task, e.g., spins around in his/her chair, sings to himself/herself, etc.
	Listening to peers talking about personal matters while interrupting learning tasks
	Listening to the teacher talk about personal matters while interrupting learning tasks
	Listening to other groups talk about personal matters while interrupting learning tasks
Organization [2]	Read through the task description
	Discuss the organization of the task with a partner
	Ask the teacher and discuss the organization of the task with the teacher
	Discuss organizational questions for the task with other groups
	Waiting for the computer, "Please wait."
	Listen to the partner talk about organizational matters
	Listen to the teacher talk about organizational matters
	Listening to other groups talk about organizational matters
Real learning time [3]	Working quietly at the computer or with paper and pencil; talking to self, mumbling, etc.
	Talking about tasks with a partner
	Asking the teacher/discussing the task with the teacher
	Talking to other groups to receive or offer help
	Watching/listening to a video (as an introduction) or watching peers and thinking with them; talking to self, mumbling, etc.
	Listening to peers talk about the topic
	Listening to the teacher talk about the topic
	Listening to other groups talk about the topic
Private - **but finished** with the learning task or task processing	Playing with cell phones or other devices while interrupting learning tasks
	Talking about personal matters with peers while interrupting learning tasks
	Talking about personal matters with the teacher while interrupting learning tasks
	Talking about personal matters with other group members while interrupting learning tasks
	Does not pay attention to the lesson and keeps to himself/herself while interrupting the learning task, e.g., spins around in his/her chair, sings to himself/herself, etc.

[4]	Listening to peers talking about personal matters while interrupting learning tasks
	Listening to the teacher talk about personal matters while interrupting learning tasks
	Listening to other groups talk about personal matters while interrupting learning tasks
[0]	Not assignable; theme changes too often during intervals

Table 2: Further examples and their classification

Characteristic	Description	Coding standards
Real learning time	Corresponds to active learning time or active time on task.	The student spends time solving the task, and this is visually or verbally recognizable. The student uses the learning program at the PC and / or comments his / her (?) own actions ("thinking aloud") regarding the problem or talks to a classmate about it. Taking part in group reflection and working with work material can also be allocated to this category.
Organization	Part of the usable learning time. Includes organization, disciplining and testing.	It can be seen or heard that the student receives working material, organizes him- or herself in a working group with other students or asks the teacher a question.
Private	Private conversations between students are also part of the usable learning time.	It can be heard that the student has a private conversation about things which happened on the weekend or during leisure time. If it can be seen that the student is surfing the internet and does not work with the learning program, this is a private activity as well.

Notes on Time-on-Task analysis

There are two methods of time-on-tasks analysis. Researchers can calculate the different ratios between the time-on-task and other classroom activities of single students as well as the classroom mean.

On the other hand, it might be interesting to just observe the ratio between the on-task time (code 3) and the off-task time (code1) or the percentage of on task and on plus off task (code3 / (code1+code3)). In this case we ignore all organization (code2) and private when finished with the task (code4) times and observe only the amount of time used when students should be working. This is interesting, for example when analyzing the effects of an intervention on students, the intervention should not be considered ineffective when teachers spend time on organization or when students are finished with the task.

Reference

Beerenwinkel, A., & Börlin, J. (2014). Surface level: teaching time, lesson phases and types of interaction. *Quality of Instruction in Physics*, 65-79.

Bostic J, Lesseig K, Sherman M, Boston M. Classroom observation and mathematics education research. *Journal of Mathematics Teacher Education*. 2021 Feb;24:5-31.

Chaturvedi, S. R. B. H., & Shweta, R. C. (2015). Evaluation of inter-rater agreement and inter-rater reliability for observational data: an overview of concepts and methods. *Journal of the Indian Academy of Applied Psychology*, 41(3), 20-27.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and psychological measurement*, *20*(1), 37-46.

Derry SJ, Pea RD, Barron B, Engle RA, Erickson F, Goldman R, Hall R, Koschmann T, Lemke JL, Sherin MG, Sherin BL. Conducting video research in the learning sciences: Guidance on selection, analysis, technology, and ethics. The journal of the learning sciences. 2010 Jan 27;19(1):3-53.

Euler, D. & Hahn, A. (2014). *Wirtschaftsdidaktik* (3. Aufl.). *UTB Wirtschaftswissenschaften, Pädagogik, Didaktik: Bd. 2525*. Haupt Verlag.

Fischer, H. E., & Neumann, K. (2012). Video analysis as a tool for understanding science instruction. In *Science education research and practice in Europe* (pp. 115-139). Brill.

Grouven, U., Bender, R., Ziegler, A., & Lange, S. (2007). Der kappa-koeffizient. *DMW-Deutsche Medizinische Wochenschrift*, *132*(S 01), e65-e68.

Hesse, H. G. (1994). *Lehr-Lern-Zeit und Lernerfolg aus psychologischer Sicht* (pp. 143-161).

Klette K. Challenges in strategies for complexity reduction in video studies. Experiences from the PISA+ study: A video study of teaching and learning in Norway. The power of video studies in investigating teaching and learning in the classroom. 2009:61-82.

Klette, K. (2023). Classroom observation as a means of understanding teaching quality: towards a shared language of teaching?. *Journal of curriculum studies*, *55*(1), 49-62.

Klette, K., & Blikstad-Balas, M. (2018). Observation manuals as lenses to classroom teaching: Pitfalls and possibilities. *European Educational Research Journal*, *17*(1), 129-146.

Pianta RC, La Paro KM, Hamre BK. Classroom Assessment Scoring System™: Manual K-3. Paul H. Brookes Publishing Co.; 2008.

Schoenfeld, A. H. (2013). Classroom observations in theory and practice. *ZDM*, *45*, 607-621.

Scholkmann, A., Siemon, J., Boom, K. D., & Knigge, M. (2017). Lernzeitnutzung im Planspielunterricht: eine Analyse des Einflusses kognitiver Fähigkeiten, Zielorientierungen und Charakteristika von Lernpartnern anhand von Videodaten. *Zeitschrift für Erziehungswissenschaft*, *20*(4), 651-669.

Seidel, T. (2005). Coding manual—surface structures: organization of classroom activities. *How to run a video study. Technical report of the IPN video study, Waxmann, New York*, 79-91.

Siemon, J., Scholkmann, A., & Paulsen, T. (2018). Beschreibung von Formen lehrerseitigen Unterstützungsverhaltens im Unterricht: Eine Analyse anhand von Videodaten. *Zeitschrift fur Bildungsforschung*, *8*(1), 19-41.

Siemon, J., Scholkmann, A., Boom, K. D., & Knigge, M. (2015). Time on Task (TT): A Manual for Coding of Student Learning Behaviour in Videos.

Stigler, J. W., & Hiebert, J. (1997). Understanding and improving classroom mathmatics instruction: An overview of the TIMSS video study. *Phi delta kappan*, *79*(1), 14.

Licensing Terms

Enqi Fan, Faculty of Education, University of Hamburg
Email: enqi.fan@uni-hamburg.de

Jens Siemon, Faculty of Education, University of Hamburg
Email: jens.siemon@uni-hamburg.de