The Impact of Instructor Presence Formats on Learning Outcomes, Visual Attention, and Cognitive Load in Educational Videos: An Eye-Tracking Study

Yen Ying Ng
Nicolaus Copernicus University in Toruń
Toruń, Poland *nyysang@hotmail.com*

Marek Placiński
Nicolaus Copernicus University in Toruń
Toruń, Poland *marpla@umk.pl*

Anna Szala
Nicolaus Copernicus University in Toruń
Toruń, Poland *aszala@umk.pl*

Abstract

Context: Educational videos are common in online learning environments, but the effectiveness of instructor presence remains debated. Some theories suggest it enhances social engagement and motivation, while others argue it increases cognitive load. **Objective:** This study investigates the effects of different instructor presentation formats (intermittent, continuous, and absent) on learning outcomes, visual attention, and cognitive load in Mandarin Chinese vocabulary learning using eye tracking. We aimed to provide objective evidence on how these formats influence attention and cognitive load, providing practical implications for designing effective educational videos. **Method:** Using a matched-groups design, 120 participants were randomly assigned to one of three conditions based on their initial performance in an online learning session. During the second session, participants watched instructional videos while their eye movements were tracked, with learning outcomes measured through speaking tests. **Results:** No significant learning outcome differences were found across conditions. Eye-tracking showed learners in all conditions prioritized Pinyin (written phonetics) while largely ignoring the instructor. **Conclusion:** These findings challenge assumptions about instructor presence and demonstrate learners' strategic visual attention regulation. Given equivalent outcomes across conditions, an instructor absent approach is preferable as it simplifies video design and reduces production complexity.

Keywords: Online Learning, Intermittent Presence, Mandarin Chinese, Language Learning, Instructional Design

1. Introduction

Educational videos have become integral to various educational settings, including blended learning, flipped classrooms, and massive open online courses (MOOCs). Despite their widespread use, research continues to investigate the most effective design strategies to enhance their educational impact. One key aspect still under investigation is the role of instructor presence.

Theoretical perspectives on the impact of a visible instructor in video lessons are somewhat contradictory [2]. According to Cognitive Load Theory (CLT), cognitive load increases when extraneous demands, such as ineffective instructional methods or environmental distractions, overburden the cognitive system, potentially hindering learning and knowledge transfer [24]. From this perspective, the presence of an instructor could be detrimental, as they may serve as an extraneous element unrelated to the core learning content. Additionally, while the instruc-

tional content and instructor may be distinct sources of information [25], instructors frequently reference on-screen content through verbal explanations and gestures [14]. This interaction could lead to a phenomenon like the split attention effect, where learners struggle to divide their focus between the instructor and the content [2].

Nonetheless, learning is deeply rooted in social interactions [3]. The Cognitive-Affective-Social Theory of Learning in digital Environments (CASTLE) proposes that social cues within digital learning contexts activate social schemas in long-term memory, triggering both social and parasocial processes that influence the cognitive mechanisms essential for effective learning [22]. Consequently, incorporating a visible instructor in educational videos may enhance learning outcomes compared to instructor absent alternatives. When learners can see an instructor, they receive multiple social cues, including eye contact, facial expressions, and gestures, that can enhanced their perceived affective and motivational engagement [2]. These cues also foster a sense of social connection, which may subsequently improve learning performance [17].

Eye-tracking methodology has been extensively employed in educational research, such as to investigate instructional design effectiveness [12]. This approach provides precise temporal and spatial data collection in real time [6], revealing attentional processes occurring during learning [11]. Specifically, it enables researchers to pinpoint exactly where attention is directed by revealing not only what people focus on but also what they neglect to observe [9]. When applied to research on instructor presence, these eye-tracking insights can elucidate whether an instructor functions as a beneficial entity that enhances engagement or as a distracting element that increases cognitive load.

Researchers have extensively investigated the effects of instructor presence in educational videos across multiple dimensions, including learning outcomes, visual attention, cognitive load, and affective variables. Despite substantial research, the field lacks consensus, with studies reporting impacts ranging from positive to negligible or even detrimental (see [18] for a detailed review). Analysis of the literature reveals several patterns. Regarding learning outcomes, instructor presence tends to either enhance performance or produce no significant difference compared to instructor absence. In contrast, findings on perceived cognitive load demonstrate considerable inconsistency across studies. The concept of perceived social presence remains relatively unexplored in this context, with current studies failing to detect statistically significant effects. These observations corroborate the conclusions drawn in Polat's [20] systematic review of the field. Furthermore, an instructor in educational videos consistently diverts learners' visual attention. This finding aligns with meta-analytic evidence from Beege et al. [2]. Although research consistently confirms this attentional diversion, the evidence indicates that this attentional shift does not necessarily translate into enhanced learning performance [19].

Several factors may contribute to the inconclusive nature of these results. First, instructors are presented in various ways [27]. Instructors may appear in different corners of the screen, and visual formats range from talking head shots to half-body presentations. Second, the instructional domain likely moderates the impact of instructor presence. This is supported by Beege et al.'s [2] meta-analysis, which identifies subject area as a significant moderator of the instructor presence effect. Third, the lack of statistically significant differences in many studies may be attributed to inherent trade-offs associated with instructor presence. Continuous instructor presence enhances social connection but potentially increases distraction, while instructor absence eliminates this distraction but sacrifices social presence benefits [13]. These competing effects may counterbalance each other [15], resulting in negligible net impacts on learning outcomes.

Building on this concern, Kizilcec et al. [13] proposed an alternative design strategy based on the hypothesis that social presence and cognitive load must be carefully balanced to achieve optimal learning outcomes. Their approach strategically varies instructor visibility: instructors appear during segments without critical learning content but are removed when important material is displayed on screen. This intermittent presentation method aims to preserve the benefits of

social presence while minimizing visual competition during key learning moments. To the best of our knowledge, only four studies have explored the potential impact of intermittent instructor presentation [13], [27], [28], [15], and none of them has employed eye-tracking methodology to compare the impacts of intermittent, continuous, and absent instructor presentation.

Moreover, within the specific domain of foreign language vocabulary learning, the effects of instructor presence may be particularly nuanced. Research indicates that visual cues such as mouth movements and hand gestures can significantly enhance speech perception and comprehension in foreign language learning [8], [16], [23]. However, the cognitive demands of simultaneously processing unfamiliar vocabulary items and phonetic patterns while attending to an instructor's visual presence may create additional cognitive burdens. This specialized vocabulary learning context warrants dedicated investigation, as the balance between multimodal communication benefits and cognitive load considerations may differ significantly from other instructional domains. Given the current gap in the literature, we conducted an eye-tracking experiment guided by the following research questions:

- **RQ1:** Does the type of instructor presence in Mandarin Chinese vocabulary learning educational videos affect learning outcomes?
- **RQ2:** Does the type of instructor presence in Mandarin Chinese vocabulary learning educational videos affect visual attention?
- **RQ3:** Does the type of instructor presence in Mandarin Chinese vocabulary learning educational videos affect cognitive load?

Our findings make a twofold contribution to research and practice in multimedia learning. First, we provide novel empirical evidence on the effects of intermittent instructor presence compared to continuous and absent conditions in Mandarin Chinese vocabulary learning, an area previously underexplored using eye-tracking methodology. Specifically, our study illuminates how learners distribute visual attention across instructional elements under varying instructor presence conditions. Second, our findings yield practical design recommendations for educational videos targeting short-term foreign vocabulary learning. Our results suggest that an instructor absent approach may be the best design strategy, as it maintains learning effectiveness while reducing production complexity and resource requirements compared to either intermittent or continuous instructor presence modes.

The rest of the paper is structured as follows: Section 2 reviews prior research investigating different instructor presentation modes in educational videos, including continuous presence, intermittent presence, and complete absence conditions, highlighting the contradictory findings and research gaps that motivate our study. Section 3 details our methodology, including experimental design, participant characteristics, materials, measurement techniques, data analysis procedures, and experiment procedure. Section 4 presents our results, systematically addressing each research question regarding learning outcomes, visual attention patterns, and cognitive load across the three instructor presence conditions. Section 5 discusses these findings, interpreting their significance in relation to established theories and prior research, particularly focusing on observed attention patterns and the absence of impact on vocabulary acquisition. Section 6 acknowledges threats to validity. Section 7 concludes with key findings, practical implications for educational video design, and future research directions in this domain.

2. Related Work

Kizilcec et al. [13] pioneered research comparing continuous and intermittent instructor presence in educational videos through a large-scale field study. They found no significant differences in learning outcomes or attrition between groups, but participants in the intermittent group reported higher cognitive load and social presence. They concluded that intermittent presence increased the instructor's visual salience, enhancing social presence while creating distraction.

Contrary to Kizilcec et al. [13], Yi et al. [27] found intermittent instructor presence superior to continuous presence in a controlled experiment. Participants in the intermittent presence group showed significantly better learning performance, lower cognitive load, and higher satisfaction, though no difference in social presence was observed. They attributed these contradictory findings to differences in experimental settings, knowledge types, and culture.

Yu [28] corroborated Yi et al. [27], finding that intermittent instructor presence significantly improved academic achievement and increased intrinsic cognitive load. Nevertheless, it was unclear whether the intermittent presentation design aligned with Kizilcec et al. [13]. Regarding subjective opinions on presentation styles, participants responded that continuous presence induced fatigue and stress, while intermittent presence enhanced curiosity and relaxation.

Kühl et al. [15] expanded the previous research by comparing how intermittent, continuous, and absent instructor presentations affect learning performance, perceived cognitive load, and social presence across two online studies. They found only a weak positive effect of intermittent presence on transfer learning in Experiment 1, which was not replicated in Experiment 2. No significant differences emerged in perceived cognitive load or social presence across conditions. Overall, their findings did not show any distinct advantages for intermittent instructor presence in online learning.

While previous studies have offered valuable insights into the effects of intermittent instructor presence, their findings remain inconsistent and are limited by their reliance on self-reported measures without eye-tracking technology. To address this gap, our study integrated eye tracking to examine how different instructor presentation formats (intermittent, continuous, and absent) influence attentional processes and cognitive load during learning. By analyzing participants' gaze patterns, we provide objective evidence on how these formats affect visual attention allocation, cognitive load, and learning outcomes. This methodological advancement offers a more nuanced understanding of the cognitive mechanisms underlying instructor presence effects, contributing both empirical data and practical implications for educational video design.

3. Methods

3.1. Experimental Design

The study employed a matched-groups design to control for potential extraneous variables that could confound the effects of instructor presentation format. The experiment consisted of two sessions. The first session, conducted online, assessed participants' language aptitude to ensure balanced group assignment based on individual performance. This facilitated their allocation into three comparable groups (intermittent, continuous, and absent instructor presence) for the second session. The second session was conducted on-site using an eye tracker, as detailed in Section 3.6. Ethical approval for the study was granted by the Scientific Research Ethics Committee of the University of Warmia and Mazury in Olsztyn.

3.2. Participants

A total of 150 participants were recruited from a Polish university. Of these, 30 were excluded: 14 did not complete the first session, and 16 withdrew afterward. The final sample consisted of 120 participants ($M_{\text{age}} = 22.66$, $SD = 4.02$; female = 85), all of whom received a monetary compensation of 80 PLN. All participants had normal or corrected-to-normal vision. They were assigned to one of three conditions for Session 2 based on their learning performance from Session 1 to ensure balanced groups: intermittent instructor presence (n = 42), continuous instructor presence (n = 40), or instructor absence (n = 38). For eye-tracking analyses, one participant was excluded due to severe data loss, resulting in a final sample of 119 ($M_{\text{age}} = 22.66$, $SD = 4.04$; female = 84). Thus, 120 participants were included in the learning outcomes analysis, while 119 were analyzed for eye-tracking data.

3.3. Apparatus and Materials

Video Lessons

One instructional video was used in Session 1 to introduce 12 Mandarin Chinese lexical items related to numbers. The 18-minute video featured two instructor presentation modes: intermittent instructor presence, where the instructor appeared only when no critical learning content was displayed, and continuous instructor presence, where the instructor's talking head was always visible in the bottom right-hand corner of the screen.

In Session 2, three videos were used in the eye-tracking experiment. Each video contained identical slides and audio narration but differed in the instructor's presentation: intermittent and continuous instructor presence following the same formats as in Session 1, plus a third condition (instructor absence) where only slides and narration were presented without any instructor. Each video in this session lasted approximately 20 minutes and was segmented into 21 trials, including an introduction, 18 vocabulary segments, a revision, and an ending. A drift correction was performed between each trial to maintain eye-tracking accuracy.

Demographic Questionnaire

The demographic questionnaire collected data on gender, age, native language, and self-reported Mandarin Chinese proficiency to control for potential confounding variables.

Knowledge Retention Exercise

The drag-and-drop tasks served as retention exercises before the learning performance test. They consist of short audio clips and pictures featuring the target vocabulary taught in each session, where participants needed to match the pronunciations with their corresponding visual representations by dragging the correct images to the appropriate audio clips. Completion of these exercises was mandatory before proceeding to the speaking tests.

Learning Performance Test

Speaking tests were conducted to evaluate participants' vocabulary retention. Participants recorded the pronunciations of as many recalled lexical items as possible.

Eye-tracking Equipment

The study utilized the Desktop Mount EyeLink 1000 Plus eye tracker, which typically offers an accuracy range of 0.25° to 0.5° (SR Research Ltd., Canada). Stimuli were presented on a 21-inch Dell monitor (1920 × 1080 pixels, 60 Hz refresh rate), with participants seated 60 cm from the screen.

3.4. Measurements

Learning Outcomes

Speaking tests assessed participants' recall of vocabulary items. Tonal accuracy was not required for correctness. One point was awarded for each correctly pronounced item (maximum: 12 points for Session 1, 18 for Session 2). Session 1 results were used to match participants into groups, while Session 2 scores served as a dependent variable.

Eye Movement Data

Eye movements were recorded monocularly at a sampling rate of 1000 Hz. A 9-point calibration and validation procedure was conducted for every participant to ensure measurement accuracy, maintaining a maximum error tolerance of 1°. If this threshold was exceeded, the procedure was repeated.

Four areas of interest (AOIs) were defined in the vocabulary learning videos: the image AOI (430,882 pixels), which displayed the picture representing the Mandarin Chinese word; the

English AOI (30,530 pixels), which contained the English translation; the Pinyin AOI (30,530 pixels), which presented the transliteration; and the instructor AOI (223,585 pixels), present only in the continuous presence condition, showing the instructor's talking head (see Figure 1).

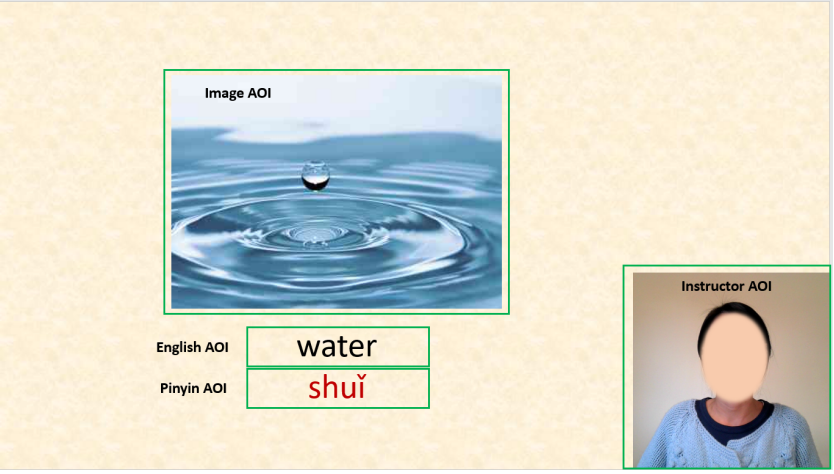

Fig. 1. Layout of AOIs in the vocabulary segment with continuous instructor presence. Note that the instructor AOI was not present in the intermittent and absent conditions.

To assess visual attention, we analyzed two key metrics: dwell time and time to first fixation. Dwell time refers to the sum of all fixation times in an AOI before the first exit [10]. It is an index that correlates with how interesting or informative an object is [5], but it also signifies difficulty in extracting information [29]. Given the variation in AOI sizes, we normalized total dwell time by dividing it by the pixel area of each AOI to allow for meaningful comparisons. Time to first fixation, on the other hand, is a latency measure that captures the difference in time between the onset of the stimulus and the first fixation on the stimulus [10]. Objects out of context and written words that are associated by speech attract early fixations, while saliency of the object is irrelevant for this measure [10].

To measure cognitive load, we consider two indices: first saccade amplitude and maximum pupil size. The former is inversely correlated with cognitive load: the more cognitively demanding the task, the shorter the saccadic amplitude [10], while the latter is positively correlated: the larger the pupil size, the more demanding the task [7].

3.5. Data Analysis

The aim of our modelling was to evaluate the influence of instructor presentation mode on participants' learning performance, visual attention, and cognitive load. We fitted separate linear mixed models (LMMs) to answer the question about learning performance and to specific ocular indices of attention and cognitive load. If possible, the models included random intercepts for participant id to ensure a better model fit [1]. Such a structure of random effects was to ensure that the observed significance and effect sizes are a result of experimental manipulation and not of individual variation [1]. In cases where there were convergence issues (a failure of the model to find an optimal and reliable solution, see [4], or singular fits (a situation in which predictors are collinear, i.e. very closely related, see [26]), we opted for a regular regression model. All analyses and visualisations were conducted using the R language (R Core Team) and its packages: lme4, tidyverse and ggplot2.

3.6. Procedure

Prior to participation, individuals received detailed information about the study and its methodology. They were also informed of their right to withdraw from the study at any point without

providing justification or incurring any consequences.

In Session 1, participants received personalized login credentials for the experimental website and provided electronic informed consent before accessing the study materials. All participants watched the same instructional video. Upon completion, they proceeded to the demographic questionnaire, followed by the knowledge retention exercise and the speaking test. The session lasted approximately 30 minutes. The data obtained from this session were analyzed to group participants based on their learning performance, after which they were invited to participate in Session 2.

Session 2 was conducted in an eye-tracking laboratory and lasted approximately 60 minutes. Participants were randomly assigned to one of three instructional conditions. Each participant was seated in front of the eye tracker, with their head stabilized using a chin and forehead rest. A nine-point calibration and validation procedure was performed prior to the experimental trials. The session followed a structured sequence: participants first viewed an introductory segment, followed by individual vocabulary segments, a revision segment, and a concluding segment. A drift correction was performed between each segment to ensure the accuracy of the eye-tracking measurements. After viewing the instructional content, participants logged into the experimental website on a researcher-provided laptop to rewatch the revision video, then completed the knowledge retention exercise and speaking test.

4. Results

We first modeled the influence of instructor presence condition on learning performance (RQ 1). The results of this analysis indicate that there is no effect of instructor presentation mode on learning when we compare continuous instructor presence with intermittent instructor presence and that there is no difference between instructor absence and intermittent instructor presence. A further pairwise comparison also does not detect a significant difference between the continuous instructor presence and instructor absence. The quantitative results are presented in Table 1 and visually represented in Figure 2.

Table 1. Effect of instructor presentation mode on learning outcomes. The intercept is the intermittent presence condition. Values rounded to two decimal places.

	Estimate	**Std. Error**	**t-value**	**p**
Intercept	6.59	0.48	13.54	<0.05
Continuous	0.31	0.69	0.44	0.66
Absent	-0.31	0.71	-0.43	0.67

We subsequently analyzed the distribution of attention (RQ2) within each condition separately, using normalized total dwell time as the first metric. The results for each condition are presented in Table 2, 3, and 4. Overall, the results indicate that participants spent more time dwelling on the English word and its Pinyin transliteration than on the image. Notably, in the continuous presence condition, the instructor's talking head did not attract more attention than the image. Post-hoc pairwise comparisons further clarify these patterns. In the intermittent presence condition, the English word attracted significantly less attention than the Pinyin transliteration (-0.07, $p < 0.01$). Similarly, in the continuous presence condition, participants spent significantly less time on the English word compared to the Pinyin transliteration (-0.05, $p < 0.01$), and the instructor attracted less attention than both the English and the Pinyin AOI (-0.01 vs -0.05 respectively, both at $p < 0.01$). The instructor absent condition showed the same pattern, with the English AOI receiving significantly less attention than the Pinyin AOI (-0.07, $p < 0.01$). In short, these results show that participants consistently paid more attention to Pinyin transliterations than to other AOIs across all experimental conditions.

The second metric used to analyze attention distribution was time to first fixation. The re-

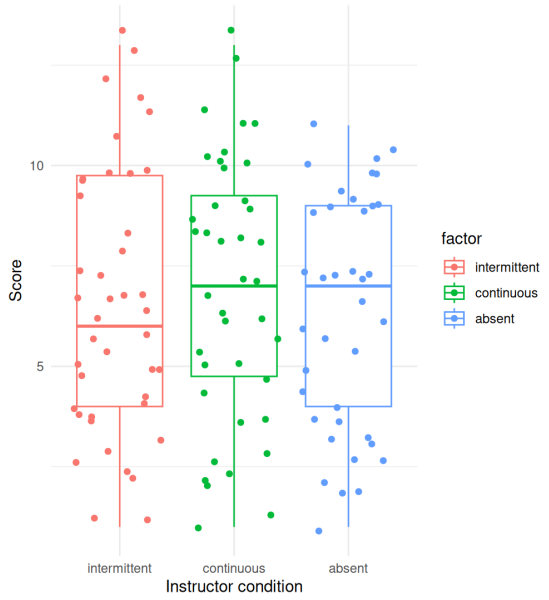

Fig. 2. Box plots representing the influence of instructor presentation mode on learning outcomes. Scattered points represent an individual participant's score.

Table 2. Normalized total dwell time differences between AOIs in the intermittent condition, with image AOI as intercept.

	Estimate	Std. Error	df	t-value	p
Intercept	4.255e-03	2.858e-03	5.517e+01	1.49	0.14
IA: English	2.463e-02	1.840e-03	2.224e+03	13.38	<0.01
IA: Pinyin	9.326e-02	1.840e-03	2.224e+03	50.68	<0.01

Table 3. Normalized total dwell time differences between AOIs in the continuous condition, with image AOI as intercept.

	Estimate	Std. Error	df	t-value	p
Intercept	4.771e-03	2.187e-03	5.959e+01	2.18	0.03
IA: Instructor	1.051e-04	1.562e-03	2.837e+03	0.07	0.94
IA: English	1.741e-02	1.562e-03	2.837e+03	11.14	<0.01
IA: Pinyin	7.094e-02	1.562e-03	2.837e+03	45.42	<0.01

Table 4. Normalized total dwell time differences between AOIs in the absent condition, with image AOI as intercept.

	Estimate	Std. Error	df	t-value	p
Intercept	4.906e-03	3.059e-03	5.162e+01	1.60	0.11
IA: English	2.484e-02	2.154e-03	1.959e+03	11.53	<0.01
IA: Pinyin	9.703e-02	2.154e-03	1.959e+03	45.05	<0.01

sults for this measure are presented in Table 5, 6, and 7for each condition. The results show that in the intermittent presence condition, participants fixated on the English AOI significantly later than on the image AOI, which in turn was fixated on significantly later than the Pinyin AOI. A similar pattern emerged in the instructor absent condition: the English AOI was fixated on significantly later than on the image AOI, and although the Pinyin AOI received earlier fixations than the image AOI, this difference did not reach statistical significance. In contrast, the continuous presence condition revealed a different pattern. Here, the image AOI attracted significantly earlier fixations than all other AOIs. Pairwise comparisons further clarify these

effects: the instructor AOI received significant later fixations than both the English AOI (1866, $p < 0.01$) and the Pinyin AOI (2131, $p < 0.01$). Additionally, the English AOI was fixated on significantly later than the Pinyin AOI (264, $p < 0.01$).

Table 5. Time to first fixation on AOIs in the intermittent condition, with image AOI as intercept.

	Estimate	Std. Error	df	t-value	p
Intercept	956.38	62.70	160.91	15.25	<0.01
IA: English	238.18	78.35	2038.51	3.04	<0.01
IA: Pinyin	-198.77	75.22	2033.84	-2.64	<0.01

Table 6. Time to first fixation on AOIs in the continuous condition, with image AOI as intercept.

	Estimate	Std. Error	df	t-value	p
Intercept	35.92	54.59	160.36	0.66	0.511
IA: Instructor	3062.30	77.26	2434.90	39.64	<0.01
IA: English	1196.13	69.33	2415.78	17.25	<0.01
IA: Pinyin	931.67	66.47	2405.09	14.02	<0.01

Table 7. Time to first fixation on AOIs in the absent condition, with image AOI as intercept.

	Estimate	Std. Error	df	t-value	p
Intercept	863.06	70.12	96.84	12.308	<0.01
IA: English	321.62	80.08	1800.68	4.016	<0.01
IA: Pinyin	-47.43	76.97	1792.20	-0.616	0.54

Finally, to evaluate the effects of instructor presence mode on cognitive load (RQ 3), we modeled the relationship between the experimental conditions and two eye-tracking metrics: first saccade amplitude and maximum pupil size. Table 8 and 9 present the results of these metrics, respectively. First saccade amplitude was significantly greater in the continuous presence condition compared to both intermittent presence and absence conditions, which showed no significant differences between them. Maximum pupil size, however, revealed no significant differences across all conditions. Further pairwise comparisons confirmed no significant difference in maximum pupil size between continuous instructor presence and instructor absence. These findings suggest that, at least to some degree, the continuous presence of an instructor may be associated with lower cognitive load, though this conclusion is supported by only one of our two eye-tracking measures.

Table 8. First saccade amplitude across conditions, with intermittent presence as the intercept.

	Estimate	Std. Error	df	t-value	p
Intercept	4.65	0.09	46.96	12.31	<0.01
Continuous	2.19	0.15	15.35	4.02	<0.01
Absent	-0.01	0.14	-0.11	-0.62	0.91

Table 9. Maximum pupil size across conditions, with intermittent presence as the intercept.

	Estimate	Std. Error	df	t-value	p
Intercept	1517.08	59.52	116.04	25.49	<0.01
Continuous	-53.44	85.20	115.94	-0.63	0.53
Absent	-101.42	86.97	116.04	-1.17	0.25

5. Discussion

The present study examined the influence of varying instructor presentation modes (intermittent, continuous, and absent) on learning outcomes, visual attention allocation, and cognitive load in Mandarin Chinese vocabulary learning. Beyond assessing learning performance across these three instructor presentation conditions, this study represents the first systematic investigation to elucidate how these distinct types of instructor presence influenced participants' visual attention patterns and experienced cognitive load in the context of foreign vocabulary learning.

Our findings revealed no statistically significant differences in learning outcomes across the three instructor presence conditions. These results corroborate previous research [15], [25], which similarly found no significant relationship between instructor presence and learning outcomes. This consistent pattern suggests that the mere visual presence of an instructor, whether continuous or intermittent, does not exert a beneficial or detrimental effects on learning. In other words, our findings do not support predictions derived from the CLT [24], the CASTLE [22], or the strategic visibility hypothesis proposed by Kizilcec et al. [13]. A compelling explanation for these results may be that participants did not perceive the instructor's visual presence as contributing substantial instructional value in this specific vocabulary learning context, consequently disregarding it during the learning phase. This interpretation is supported by our eye-tracking data, which demonstrated that participants allocated minimal visual attention to the instructor during the learning process.

Across all conditions, participants exhibited significantly greater total dwell time on the Pinyin AOI compared to other AOIs, indicating its high informational value and demonstrating a prioritization of phonetic information during foreign language vocabulary learning. Notably, contrary to previous findings regarding the attention-capturing nature of instructors in educational videos [14], [25], the instructor's talking head in the continuous presence condition did not attract a substantial amount of visual attention. Instead, it received comparable attention to the image AOI, suggesting that the instructor did not function as a significant visual attractor despite its dynamic qualities. Furthermore, participants fixated on the Pinyin transliteration significantly faster than on either the corresponding English translation or the instructor. In the continuous presence condition, the instructor AOI received significantly delayed fixations relative to all other AOIs. This temporal lag in visual engagement with the instructor suggests that participants perceived the instructor's visual presence as peripherally relevant to the immediate vocabulary learning objectives. That is, they attended to the instructor only after processing the essential information on the screen. These findings collectively suggest that participants demonstrated strategic visual attention allocation, effectively identifying and prioritizing the most critical visual elements for foreign vocabulary learning. This selective attentional focus challenges previous concerns about instructor presence potentially inducing a kind of split attention effect in multimedia learning environments [2]. Instead, our results indicate that learners in this specific context were able to regulate their visual attention to optimize information processing, effectively mitigating any potential cognitive overload from the instructor's presence.

Previous research on the influence of visual cues in speech perception and comprehension within foreign language learning contexts has consistently reported positive outcomes [8], [16], [23]. Nonetheless, our study did not replicate these findings, likely due to two key differences in instructional design. First, the video materials used in these studies predominantly featured an instructor alone, without simultaneous presentation of learning content. Second, the instructors in those studies deliberately incorporated hand gestures and pronounced mouth movements during instruction. These intentional non-verbal communicative elements were specifically designed to enhance linguistic information processing, whereas the instructor in our study maintained a more neutral presentation style without emphasizing such pedagogical visual cues, particularly hand gestures. This fundamental difference in instructional design may explain the discrepancy between our findings and those of previous research regarding the facilitative roles

of instructor presence in language learning.

Lastly, the analysis of cognitive load through eye-tracking metrics yielded mixed results, providing partial support for the impact of instructor presence mode on learners' cognitive processing. First saccade amplitude was significantly larger in the continuous presence condition compared to the other conditions, suggesting that continuous instructor presence may help reduce cognitive load. However, maximum pupil size did not differ significantly across the three conditions. A plausible explanation for this discrepancy is that cognitive and emotional influences on pupil diameter are relatively subtle [10], and thus, the fluctuations in mental workload in this study was not substantial enough to be detected through this metric. These inconsistent findings underscore the need for caution when interpreting cognitive load solely based on eye-tracking metrics. Taken together, the results provide partial support for the hypothesis that continuous instructor presence can alleviate cognitive load during learning. However, the absence of converging evidence from both eye-tracking metrics cautions against a definitive conclusion.

6. Threats to Validity

While this study provides valuable insights into the effects of instructor presence on Mandarin vocabulary learning, several threats to validity must be acknowledged and considered when interpreting the findings.

Several threats to *construct validity* are present in this study. The assessment of learning outcomes primarily measured short-term knowledge retention without evaluating long-term knowledge retention and knowledge transfer, which may not fully verify effective learning. Moreover, while eye-tracking metrics provide objective measures of visual attention, they may not fully reflect the qualitative aspects of attention, such as the reasons behind specific fixations. Finally, the measurements for cognitive load serve as indirect indicators of mental effort and can be influenced by factors unrelated to cognitive processing, such as lighting conditions, emotional arousal, or individual differences in physiological responses.

The eye-tracking laboratory setting presents a threat to *internal validity*. Participants' awareness of being monitored may have induced artificial viewing patterns that deviate from authentic learning behaviors. For instance, participants could have exhibited heightened focus on specific AOIs which they perceived as "crucial" for the study, rather than engaging in natural viewing behaviors.

Two threats to *external validity* limit the generalizability of the findings. The first one is the representativeness of participants. This is because our participant pool consisted predominantly of Polish young adult university students, which limits generalizability to other age groups, educational backgrounds, and cultural contexts. The second threat is related with the learning context. In this experiment, participants viewed only one video lesson per session before completing posttests, which inadequately represents typical online learning environments where courses span weeks or months and include multiple lessons before assessment. Furthermore, to control for extraneous variables, videos were presented only once without replay options, which did not fully simulate authentic learning experiences where students typically control their learning pace and can review materials as needed.

7. Conclusion

This study investigated the impact of instructor presentation modes on Mandarin vocabulary learning. Our findings revealed no significant differences in learning outcomes across continuous, intermittent, or absent instructor conditions. Eye-tracking data showed learners consistently prioritized written phonetic information (Pinyin) while allocating minimal attention to the instructor when present. These results challenge assumptions regarding both beneficial and detrimental effects of instructor visibility in educational videos and suggest learners effectively

regulate visual attention to optimize information processing. The consistent prioritization of Pinyin highlights the critical role of transliteration in foreign language vocabulary learning.

From a practical perspective, our findings suggest that an instructor absent approach may represent an efficient design strategy for vocabulary learning videos, provided that written phonetic information is clearly displayed on the screen. This approach maintains pedagogical effectiveness while reducing production complexity and resources.

Future research should explore these findings across diverse populations, including younger learners, older adults, and various cultural groups, to determine whether visual attention patterns and learning outcomes differ. Studies conducted in more naturalistic learning environments, where learners can control video playback and engage with multiple lessons over extended periods, would better reflect authentic online learning experiences. Additionally, investigating the effects of instructor visibility on long-term retention and application of vocabulary knowledge would provide valuable insights into the sustainability and transferability of learning outcomes. Finally, future work could explore whether immersive technologies such as virtual reality [21] can offer a more effective form of instructor presence.

Acknowledgements

This research was funded by the state budget under the Pearls of Science program of the Polish Ministry of Science and Higher Education (Project No. PN/01/0232/2022), with a funding amount and total project value of 142164 PLN.

References

[1] Barr, D., Levy, R., Scheepers, C., and Tily, H.: Random Effects Structure for Confirmatory Hypothesis Testing: Keep It Maximal. In: *J. Mem. Lang.* 68.3 (2013), pp. 255–278.

[2] Beege, M., Schroeder, N. L., Heidig, S., Rey, G. D., and Schneider, S.: The Instructor Presence Effect and Its Moderators in Instructional Video: A Series of Meta-analyses. In: *Educ. Res. Rev.* 41 (2023).

[3] De Felice, S., Hamilton, A. F. d. C., Ponari, M., and Vigliocco, G.: Learning from Others is Good, with Others is Better: The Role of Social Interaction in Human Acquisition of New Knowledge. In: *Philos. Trans. R. Soc. B.* 378.1870 (2023).

[4] Eager, C. and Roy, J.: *Mixed Effects Models are Sometimes Terrible*. 2017. arXiv: 1701.04858 [stat.AP]. URL: https://arxiv.org/abs/1701.04858.

[5] Friedman, A. and Liebelt, L.: On the Time Course of Viewing Pictures with a View Towards Remembering. In: *Eye Movements: Cognitive and Visual Perception*. Ed. by Fisher, D., Monty, R., and Senders, J. Lawrence Erlbaum Associates, 1981, pp. 137–155.

[6] Godfroid, A.: Eye Tracking in Second Language Acquisition and Bilingualism: A Research Synthesis and Methodological Guide. 1st ed. Routledge, 2020.

[7] Hess, E. H. and Polt, J. M.: Pupil Size in Relation to Mental Activity during Simple Problem-Solving. In: *Science* 143.3611 (1964), pp. 1190–1192.

[8] Hirata, Y. and Kelly, S.: Effects of Lips and Hands on Auditory Learning of Second-Language Speech Sounds. In: *J. Speech Lang. Hear. Res.* 53 (2010), pp. 298–310.

[9] Hogan, M., Barry, C., and Lang, M.: Dissecting Optional Micro-Decisions in Online Transactions: Perceptions, Deceptions, and Errors. In: *ACM Trans. Comput.-Hum. Interact.* 29.6 (2022), pp. 1–27.

[10] Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., and van de Weijer, J.: Eye Tracking : A Comprehensive Guide to Methods and Measures. Oxford University Press, 2011.

[11] Hyönä, J.: The Use of Eye Movements in the Study of Multimedia Learning. In: *Learn. Instr.* 20.2 (2010), pp. 172–176.

[12] Jarodzka, H., Holmqvist, K., and Gruber, H.: Eye Tracking in Educational Science: Theoretical Frameworks and Research Agendas. In: *J. Eye Mov. Res.* 10 (2017), pp. 1–18.

[13] Kizilcec, R. F., Bailenson, J. N., and Gomez, C. J.: The Instructor's Face in Video Instruction: Evidence from Two Large-scale Field Studies. In: *J. Educ. Psychol.* 107.3 (2015), pp. 724–739.

[14] Kizilcec, R. F., Papadopoulos, K., and Sritanyaratana, L.: Showing Face in Video Instruction: Effects on Information Retention, Visual Attention, and Affect. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '14. Toronto, Ontario, Canada: Association for Computing Machinery, 2014, pp. 2095–2102.

[15] Kühl, T., Teske, F., Merkt, M., and Sondermann, C.: Temporary Instructor Presence in Lecture Slides Does Not Enhance Online Learning. In: *Educ. Technol. Res. Dev.* (2024).

[16] Kushch, O., Igualada, A., and Prieto, P.: Prominence in Speech and Gesture Favour Second Language Novel Word Learning. In: *Lang. Cogn. Neurosci.* 33 (2018), pp. 992–1004.

[17] Mayer, R. E.: Multimedia Learning. 3rd ed. Cambridge University Press, 2020.

[18] Ng, Y. Y. and Przybylek, A.: Instructor Presence in Video Lectures: Preliminary Findings From an Online Experiment. In: *IEEE Access.* 9 (2021), pp. 36485–36499.

[19] Pi, Z., Ling, H., Li, X., and Wang, Q.: Instructors' Pointing Gestures and Positive Facial Expressions Hinder Learning in Video Lectures: Insights from Teachers and Students in China. In: *Educ. Inf. Technol.* 29.16 (2024), pp. 21115–21131.

[20] Polat, H.: Instructors' Presence in Instructional Videos: A Systematic Review. In: *Educ. Inf. Technol.* 28 (2023), pp. 8537–8569.

[21] Przybyłek, A., Belter, D., and Conboy, K.: A study of Scrum @ S&P Global in the post-COVID-19 era: Unsuitable for remote work or just flawed implementation? In: *Information and Software Technology* 183 (2025), p. 107728.

[22] Schneider, S., Beege, M., Nebel, S., Schnaubert, L., and Rey, G. D.: The Cognitive-Affective-Social Theory of Learning in digital Environments (CASTLE). In: *Educ. Psychol. Rev.* 34 (2022), pp. 1–38.

[23] Skipper, J. I., Goldin-Meadow, S., Nusbaum, H. C., and Small, S. L.: Gestures Orchestrate Brain Networks for Language Understanding. In: *Curr. Biol.* 19.8 (2009), pp. 661–667.

[24] Sweller, J., van Merriënboer, J. J. G., and Paas, F.: Cognitive Architecture and Instructional Design: 20 Years Later. In: *Educ. Psychol. Rev.* 31.2 (2019), pp. 261–292.

[25] van Wermeskerken, M., Ravensbergen, S., and van Gog, T.: Effects of Instructor Presence in Video Modeling Examples on Attention and Learning. In: *Comput. Hum. Behav.* 89 (2018), pp. 430–438.

[26] Winter, B.: Statistics for Linguists: An Introduction Using R. 1st ed. Routledge, 2019.

[27] Yi, T., Yang, X., Pi, Z., Huang, L., and Yang, J.: Teachers' Continuous vs. Intermittent Presence in Procedural Knowledge Instructional Videos. In: *Innov. Educ. Teach. Int.* 56.4 (2019), pp. 481–492.

[28] Yu, Z.: The Effect of Teacher Presence in Videos on Intrinsic Cognitive Loads and Academic Achievements. In: *Innov. Educ. Teach. Int.* 59.5 (2021), pp. 574–585.

[29] Zagermann, J., Pfeil, U., and Reiterer, H.: Measuring Cognitive Load using Eye Tracking Technology in Visual Computing. In: *Proceedings of the Sixth Workshop on Beyond Time and Errors on Novel Evaluation Methods for Visualization*. BELIV '16. Baltimore, MD, USA: Association for Computing Machinery, 2016, pp. 78–85.