# EVALUATION OF COMPUTER VISION-AIDED MULTIMEDIA LEARNING IN CONSTRUCTION ENGINEERING EDUCATION

***Anthony O. Yusuf, Adedeji O. Afolabi, Abiola A. Akanmu, & Johnson Olayiwola***
*Myers-Lawson School of Construction, Virginia Tech, Blacksburg, VA, USA.*

***ABSTRACT:*** *Due to the practice-oriented nature of construction engineering education and barriers associated with physical site visits, videos are invaluable means to expose students to practical curricula content. Prior studies have investigated various design principles of multimedia pedagogical tools to enhance student learning and reduce cognitive load. These design principles and computer vision techniques can afford the design and usage of a multimedia learning environment with annotated content to teach students construction safety practices. Hence, using subjective and objective measures such as self-reported cognitive load, eye tracking metrics and verbal feedback, this study assesses the effectiveness of a computer vision-aided multimedia learning environment as well as examines variations across students' demographics. Students were exposed to both annotated and unannotated versions of the learning environment. The annotated version of the learning environment was considered more effective in triggering students' attention to learning content, but higher cognitive load levels were reported by participants. The same demographic groups that dwelled longer and on more annotated areas of interest also reported higher overall cognitive load. Keeping with individual differences principle of multimedia learning, demographic variations in participants' cognitive load and effectiveness of the learning environment were reported. The study provides implications for instructors in construction engineering programs on effective use of computer vision-aided annotated videos as instructional materials. This study could serve as a benchmark for future studies on artificial intelligence techniques for signaling in multimedia learning. This study reveals the affordances of computer vision-aided multimedia learning in construction engineering education and the need for adaptation of multimedia learning tools to students' demographics.*

***KEYWORDS***: *Computer vision, construction engineering education, demographic differences, multimedia learning, video.*

## 1. INTRODUCTION

Construction-related disciplines are applied science; hence they are rich in practical components which are usually difficult for instructors to cover in the classroom (Gunhan, 2015). The imbalance between theory and practice has been one of the challenges in preparing students for the workplace (Afonso et al., 2012). Hence, academia is in constant effort to achieve a proper blend of theory and practice (Bozoglu, 2016). Site visits are being used to circumvent this challenge by exposing students to real-world examples, spatio-temporal scenarios of construction operations and interaction with practitioners (Eiris Pereira & Gheisari, 2019). However, barriers associated with site visits such as safety, coordination, distance/location, limited what-if scenarios, and concerns for disabled students(Eiris Pereira & Gheisari, 2019) have necessitated the need for new methods of bringing practical examples into the classroom. Videos are now increasingly being widely used as pedagogical tools to address these limitations (Shojaei et al., 2021). Videos enable instructors to bring the real world into the classroom. Videos also allow for experiments, site visits, and demonstrations that otherwise would have been impossible. Beyond knowledge transmission, videos expose students to diverse experiences, attitudes, and emotions and promote interactions and discussions (Ferreira et al., 2013). However, the use of videos also comes with some challenges. For example, if not intelligently designed, videos could be ineffective for learning because they could increase cognitive load, and not capture learners' attention (De Koning et al., 2009). In addition, videos could contain non-essential information which could be distracting to learners. These downsides of videos have been earlier reported (Homer et al., 2008). To circumvent these challenges, the adoption of multimedia learning principles such as removal of extraneous content and signaling of important learning content are effective measures (Mayer & Fiorella, 2014). Signaling involves the use of cues (e.g., arrow, boundary boxes, color contrast) to point out important learning content to learners in a multimedia environment. In other domains, previous studies have demonstrated the effectiveness of these techniques in ensuring that videos are effective pedagogical tools (De Koning et al., 2009; Navarro et al., 2015).

Given the advances in computing, its affordances and wide applications, manual signaling methods which could be laborious, undynamic and time intensive can be replaced with automation afforded by artificial intelligence. By leveraging computer vision (CV) techniques (such as object and interaction detection), construction videos can be automatically annotated to call out specific learning contents. This has been demonstrated in other endeavors such as detecting human daily activities using convolutional neural networks (Zhang et al., 2017). Adopting this in construction engineering education is important given the need to visualize theoretical concepts,

Referee List (DOI: 10.36253/fup_referee_list)
FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup_best_practice)

Anthony Yusuf, Adedeji Afolabi, Abiola Akanmu, Johnson Olayiwola, *Evaluation of Computer Vision-Aided Multimedia Learning in Construction Engineering Education,* pp. 242-253, © 2023 Author(s), CC BY NC 4.0, DOI 10.36253/979-12-215-0289-3.23

understand the pace and sequence of construction tasks, and spatiotemporal nature of construction activities (Eiris Pereira & Gheisari, 2019). Previous studies (Abdulrahaman et al., 2020; Stark et al., 2018) have highlighted the need to evaluate the efficacy of multimedia learning environments. To evaluate multimedia pedagogical tools, learners' cognitive load level and demographic differences have been suggested as primary considerations (Grimley, 2007). Also, earlier studies have combined objective and subjective measures and compared two or more multimedia learning environments (Abdulrahaman et al., 2020; Stark et al., 2018). Despite the potential of multimedia learning in construction engineering education, it has received little attention in literature, especially the application of CV in multimedia learning. Hence, this study compared two multimedia learning environments from the lens of demographic differences to evaluate the effectiveness of CV-aided multimedia learning in construction engineering education.

# 2. BACKGROUND

## 2.1 Application of Computer Vision in Education

Computer vision is being increasingly leveraged in several educational contexts because of its value to conventional educational methods by improving teaching and learning (Savov et al., 2018; Sophokleous et al., 2021). For instance, using a face recognition algorithm, Savov et al. (2018) combined computer vision, and internet of things to provide engaging experience for students. The study demonstrated the efficacy of computer vision through adaptation to learners' facial expression to help instructors tailor the teaching process to learners' preferences. Tetiana et al. (2021) also leveraged computer vision and augmented reality to allow students to interact with and obtain additional virtual information about research objects. This helped to promote effective interaction between students and educational material. Similarly, using facial emotions, pose estimation, and head rotation, Poonja et al. (2023) developed a computer vision-based system to detect students' engagement in online learning. Computer vision has been adopted in other educational context such as enhancing the teaching of mechatronics (Tudić et al., 2022), in distance education of new generation labor productivity (Zhao & Li, 2021), as well as educational robotics in K-12 education (Sophokleous et al., 2021). Interaction and object detection techniques of computer vision can be leveraged to signal essential learning contents in videos for teaching purposes. For example, Tang et al. (2020) used Faster Region-proposal Convolutional Neural Network (Faster RCNN) to detect workers and materials for safety monitoring. Using deep residual learning network, Hashimoto et al. (2019) used computer vision for automated operative step detection during Laparoscopic Sleeve Gastrectomy. Similarly, Aronson (2018) demonstrated the efficacy of computer vision for signaling violation of human right in videos. However, there are scarce studies that demonstrated the effectiveness of computer vision for signaling in multimedia learning.

## 2.2 Evaluation of Multimedia Learning Tools

To evaluate multimedia learning tools, a comparison approach which involves comparing two or more multimedia pedagogical tools is a common method (Abdulrahaman et al., 2020). For example, Chiu et al. (2018) compared the efficacy of annotated and unannotated versions of a video to teach cardiopulmonary resuscitation. Using pre and posttests, eye tracking metrics, satisfaction and self-reported cognitive load questionnaires, the study reported that students that learned with annotations had lower cognitive load, concentrated more on the critical parts of the instructional video, and thus learned more effectively and easily. Also, combinations of objective and subjective measures have been encouraged in usability evaluation (Abdulrahaman et al., 2020). This is due to the limitations of subjective measures such as risk of prejudice, lack of response, and lack of supporting evidence for respondents' ratings (Kelley et al., 2003). Objective measures such as eye tracking metrics (e.g., fixations and dwell times) are widely used in the usability evaluation of multimedia learning tools (Molina et al., 2018; Stark et al., 2018). National Aeronautics and Space Administration Task Load Index (NASA-TLX) is a widely used subjective evaluation tool for assessing cognitive workload. NASA TLX assessed cognitive workload across six subscales: Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, and Frustration (Sharek, 2011). Eye Tracking and NASA TLX have been used in previous studies (Latifzadeh et al., 2020; Law et al., 2010) for the evaluation of multimedia pedagogical tools. Other subjective measures such as think-aloud protocol, interview, and verbal feedback are also being used (Abdulrahaman et al., 2020). Also, to evaluate the usability of multimedia pedagogical tools, demographic differences such as gender (Grimley, 2007), academic level, academic program (Castro-Alonso et al., 2019), ethnicity (Moreno & Flowerday, 2006) and prior experience (Kalyuga et al., 2000) are deemed important considerations.

# 3. METHODOLOGY

## 3.1 Overview

The study evaluated the efficacy of an annotated video designed to teach construction students different construction safety practices. A comparison approach was adopted. Students were exposed to two learning environments. One was designed with computer-vision aided signals or cues to call out essential learning content while the other was not. Subjective and objective measures of the participants in the two learning environments were compared. In addition, keeping with individual differences principle of multimedia learning, demographic variations in participants' cognitive load and eye tracking metrics in the annotated learning environment were assessed.

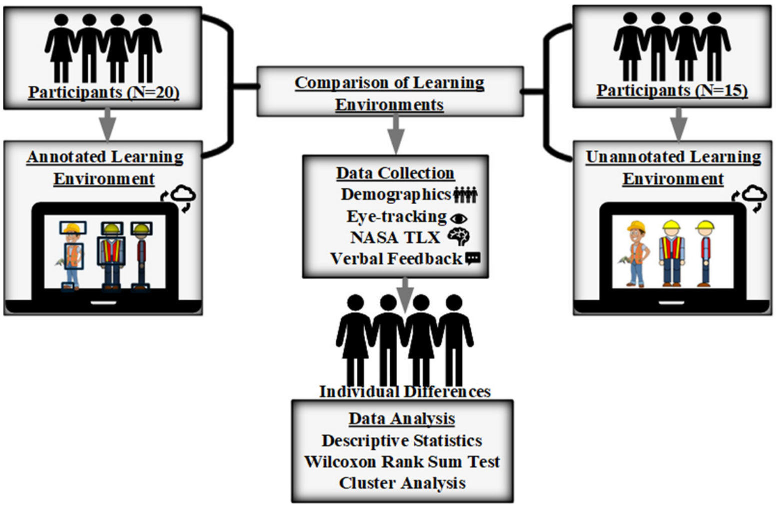


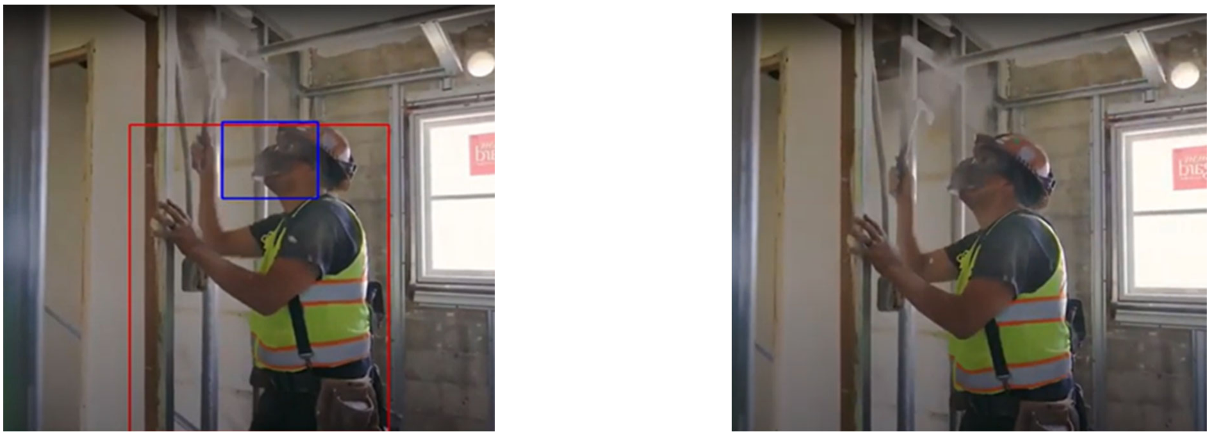
Fig. 1: Overview of Methodology

## 3.2 The Learning Environments: Annotated and Unannotated Videos

The two versions of the video had the same contents, each about 6-minutes long. The videos contain both visual and audio presentations of eleven (11) construction safety practices. These include Personal Protective Equipment (PPE), Situational awareness (SA), Secure workers at height (SWH), Securing materials (SM), Exclusion zone (EZ), Deep excavation (DE), Signage (S), Fall protection (FP), Mobile phone use (MPU), Ladder use (LU), Ergonomics (ER). These safety practices were identified by Olayiwola, Yusuf, et al. (2023) as critical to complement classroom teaching during site visits. The safety practices represent areas of interest (AOIs). Safety practices were chosen because Pedro et al. (2016) highlighted the need for knowledge of safety practices in preparing the future workforce. The annotated video contains computer vision-aided signals to call-out the safety practices while the unannotated video does not. Both object and interaction detection techniques of computer vision were used in this study. Faster RCNN, a deep learning technique was combined with Visual Geometry Group network (VGG16) (a convolution neural network architecture) to signal the construction safety practices with boundary boxes. Visual translation embedding network (VTransE) was used as the interaction detection technique. The details of the design and development of the annotated learning environment are presented in Olayiwola, Akanmu, et al. (2023). Examples of frames from the annotated and unannotated videos are shown in Figure 2.

## 3.3 Participants and study approval

After the Virginia Tech Institutional Review Board approved the study, thirty-five (35) participants who are students in construction-related programs volunteered to participate. Twenty (20) of them used the annotated learning environment while fifteen (15) used the unannotated learning environment. All the participants are between 20 to 24 years old. The participants' demographics are shown in Table 1 below.



(a) Sample frame of annotated video (b) Sample frame of unannotated video

Fig. 2: Frames of Annotated and Unannotated Videos.

Table 1: Participants' Demographic Information.

| Demographics | Experimental Group (N=20) | Control Group (N=15) |
|---|---|---|
| **Gender** | | |
| Male | 12 | 12 |
| Female | 8 | 3 |
| **Academic Program** | | |
| Building Construction (BC) | 5 | 5 |
| Construction Engineering and Management (CEM) | 9 | 10 |
| Civil and Environmental Engineering (CEE) | 6 | - |
| **Academic Level** | | |
| Junior | 4 | 10 |
| Sophomore | 16 | 5 |
| **Years of construction experience** | | |
| Less than 2 | 13 | 11 |
| 2-5 | 7 | 4 |
| **Ethnicity** | | |
| White/Caucasian | 12 | 9 |
| Asian | 5 | 3 |
| African American | 2 | 2 |
| Hispanic/Latino | 1 | 1 |

## 3.4 Experimental design and Data collection

Before the experiments commenced, every participant was intimated with the workflow of the experimental procedure. Thereafter, the participants signed the informed consent form and completed the demographic questionnaire. A web-based eye tracker (Gaze-recorder) was used in this study. The participants' eyes were calibrated, and then they watched the 6-minute video of construction safety practices. Eye-tracking data was collected as the participants watched the video. Two separate experiments were conducted. In the first experiment, the experimental group (n = 20) was exposed to the annotated video while in the second experiment, the control group (n = 15) was exposed to the unannotated video. After every session, participants completed three subscales of the NASA-TLX questionnaire (i.e., Mental demand, Effort, and Frustration). These subscales have been used to assess cognitive load in multimedia learning (Refat et al., 2020). Finally, the participants' verbal feedback was audio-recorded. Each session lasted for about one (1) hour.

### 3.5 Data analysis

Dwell times of the participants were collected for the eleven AOIs in the videos. Wilcoxon Rank Sum Test was used to test for statistically significant differences since the comparisons were between two independent groups and independent observations. Descriptive statistics were used for the self-reported cognitive load. MS Office Excel and SPSS were used for the analysis. The verbal feedback was transcribed and analyzed using cluster analysis. The analyses were done by comparing the participants’ dwell times on the AOIs of the annotated and unannotated videos and the demographic differences of the participants.

## 4. RESULTS AND DISCUSSION

### 4.1 Dwell Time Comparison in Annotated and Unannotated Video

As shown in Figure 3, significant differences in the dwell times were found between the annotated and unannotated videos for seven (7) AOIs of which four (4) had significantly higher dwell times in the annotated video. These include PPE, Situational Awareness, Securing Material, and Signage ($p < 0.05$). The overall dwell time of the participants was higher in the unannotated environment although no statistically significant difference was observed. However, the participants dwelled longer on more of the AOIs in the annotated video. The participants dwelled longer on 7 out of the 11 safety practices in the annotated video. In the unannotated video, the participants only dwelled longer on four (4) safety practices, which include Secure workers at height, Fall Protection, Use of Mobile Phone and Ergonomics. This result shows that although the participants spent more time in the unannotated environment, they did not dwell longer on more AOIs. Whereas the participants spent less time in the annotated environment but dwell longer on the AOIs. This shows that the annotation was effective to direct the learners to important learning content and to stimulate their interest. This finding agrees with Molina et al. (2018) who explained that learners might dwell more on AOIs in multimedia learning. The higher dwell times on the AOIs in the annotated video shows the extent of focus and interest in the AOIs (Bojko, 2013; Carter & Luke, 2020). This result aligns with the findings of previous studies (Molina et al., 2018; Navarro et al., 2015) which underscored the potential of signals to trigger learners’ interest, improved learners’ visual search efficiency and provide greater visibility for important learning content.

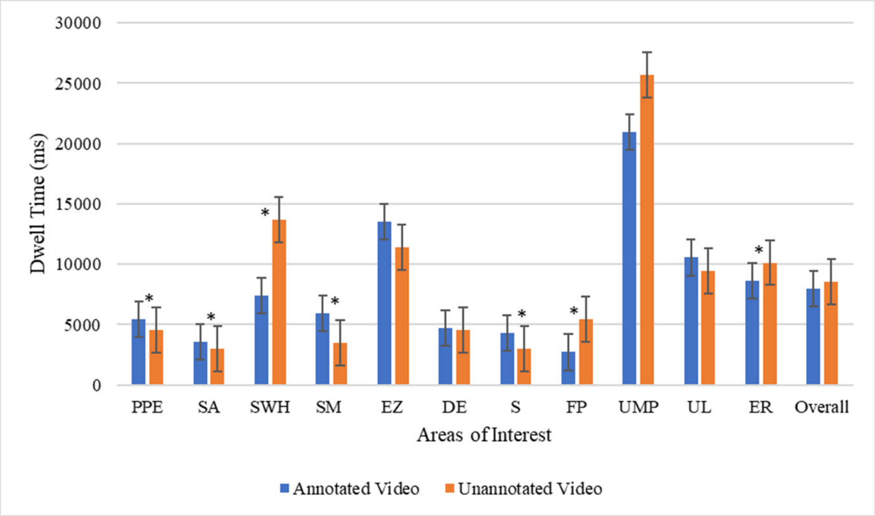


Fig. 3: Dwell Time on AOI

## 4.2 Comparison of Demographic Differences within Annotated Video

### 4.2.1 Gender

Within the annotated environment, gender differences were observed, as shown in Figure 4. On the overall, female students dwelled longer on the AOIs than their male counterparts. The overall dwell time of the female students was statistically higher ($p$ < 0.05). Also, the female participants dwelled more on each of the AOIs than their male counterparts. This phenomenon reveals that signals were efficacious in drawing the attention and stimulating the interest of female students than male students. The female participants' dwell times on the AOIs were significantly higher for three (3) AOIs, which include Secure Workers at Height, Deep Excavation and Use of Mobile Phone. This could mean that female students prefer to learn with annotated videos than male students. This could be helpful to instructors in their choice of instructional materials. This finding contributes to prior studies (Grimley, 2007; Saha & Halder, 2016) which have shown gender differences in information process in multimedia learning. The finding agrees with Dousay and Trujillo (2019) who reported that females had higher situational interest in multimedia learning than males.

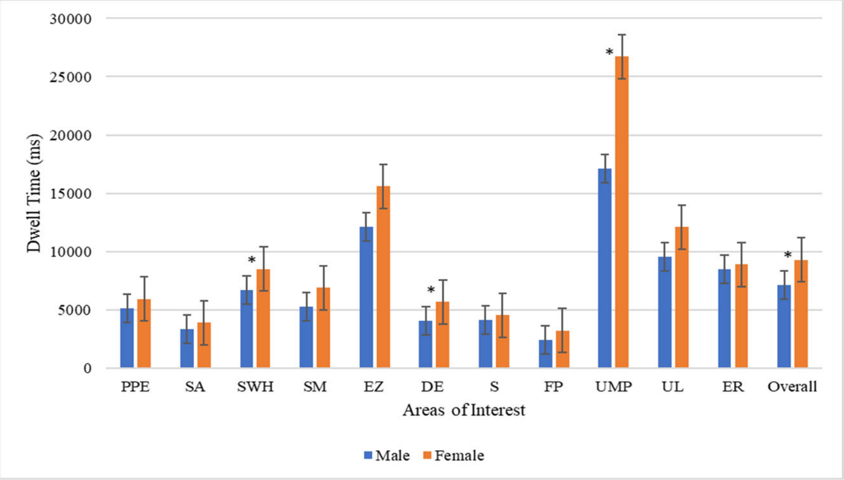


Fig. 4: Male and Female Dwell Time for Annotated Video

### 4.2.2 Academic Level

On the overall, Figure 5 shows that the senior-level students dwelled longer (although not significant, $p$>0.05) on the annotated video than junior-level students. Also, the senior-level students dwelled more on nine (9) out of the eleven (11) AOIs than their Junior-level counterparts. The Senior-level participants dwelled more on all AOIs except Use of Mobile Phone and Ergonomics. The results showed that the Senior-level participants significantly dwelled longer ($p$<0.05) on Exclusion Zone ($p$<0.05). The academic level of the students could be synonymous with prior knowledge. Although, previous studies (Grimley, 2007; Kalyuga, 2013) have noted that multimedia learning would be effective for learners with lower prior knowledge, the findings of this study differ from Navarro et al. (2015) who reported that students of lower academic levels dwelled more on AOIs while senior students only take a glance. The difference in the finding could be because the participants in this study were college students (aged 20-24 years) while those in prior studies (Grimley, 2007; Navarro et al., 2015) were primary school pupils (aged ≤ 11 years). Also, in this study, better than junior-level participants, the senior-level participants might have been more willing to explore the annotated learning environment and found the signaled concepts more engaging which might have been responsible for dwelling on more signaled concepts. This result contributes to earlier studies, e.g., Castro-Alonso et al. (2019), on academic level-based differences in multimedia learning.

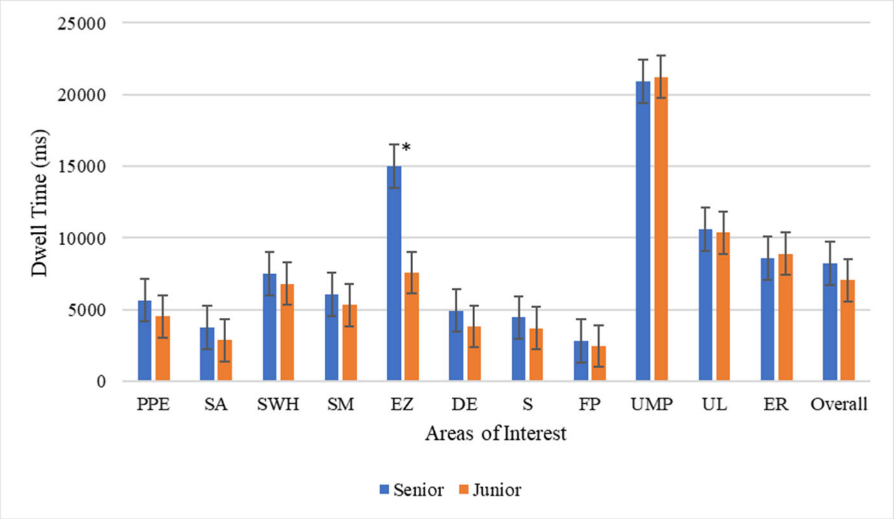


Fig. 5: Junior and Senior Dwell Time for Annotated Video

### 4.2.3 Years of Experience

As shown in Figure 6, although no statistically significant difference was observed, overall, students with 0-2 years of experience had higher dwell time on the AOIs. The students dwelled more on the AOIs (8 out of 11) than those with 2-5 years of experience. The students with 2-5 years of experience only dwelled longer on PPE, Situational Awareness and Secure of Materials. This result shows that the learners with lower experience would perceive better learning benefits with the annotated learning environment. This study contributes to prior studies e.g., Kalyuga et al. (2000) that have demonstrated differences based on prior experience in multimedia learning.

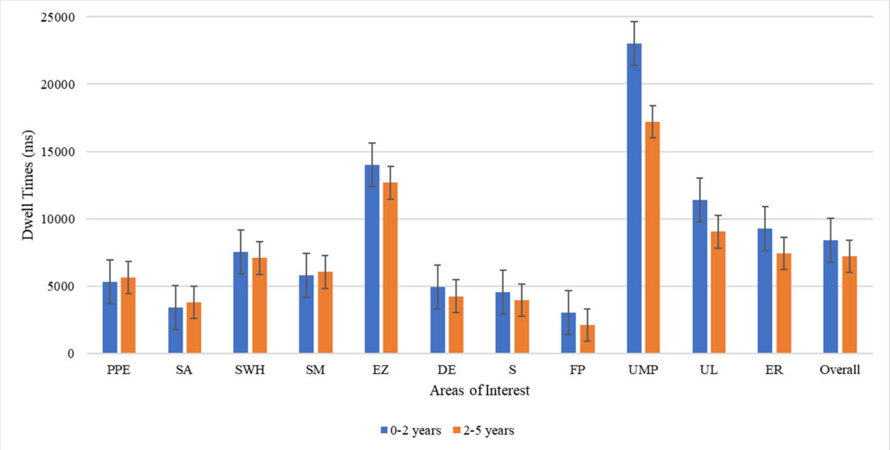


Fig. 6: Dwell Time for Annotated Video Based on Years of Experience

### 4.2.4 Academic Program

No statistically significant difference was observed in the comparison based on students' academic program. On the overall, students in CEE dwelled longer on more AOIs than those in BC and CEM respectively (Figure 7). For six (6) of the AOIs namely, Situational Awareness, Exclusion Zone, Deep Excavation, Fall Protection, Use of Mobile Phone, Use of Ladder, the ascending order of the increase in dwell time on the AOIs is BC students, CEM students and CEE students. The variation observed based on academic programs could be due to differences in the emphasis of each program even though they are all aspects of construction education. For instance, Abudayyeh et al. (2000) pointed out that the CEE program is focused more on design of facilities; CEM program has concentration on achieving a balance between the engineering and management component of construction, while BC program has emphasis on management and business components of construction. Hence, the differences in the educational background of the students could have been responsible for the variations. For example, since CEE students are in a program that focused on design of infrastructural facilities, they might have been less familiar with the construction safety practices in the annotated video compared to their counterparts in CEM and BC programs. This could account for their higher dwell times on most of the AOIs. This result contributes to earlier studies e.g., Castro-Alonso et al. (2019) on differences based on academic program in multimedia learning.

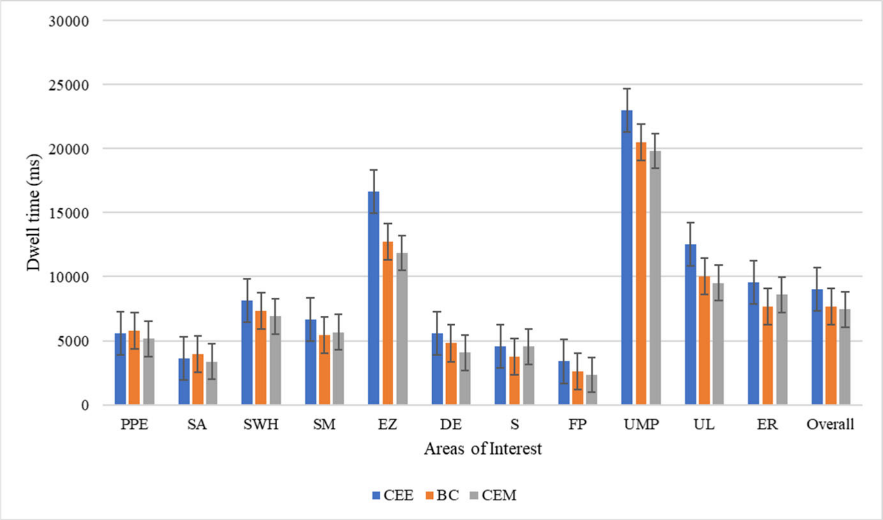


Fig. 7: Dwell Time for Annotated Video Based on Academic Program

### 4.2.5 Ethnicity

Though without any significant difference, the result reveals that on the overall, White students dwelled longer on the annotated video (Figure 8). They also dwelled longer on more AOIs (6 out of 11) than students of other ethnicities. Studies comparing ethnic differences in multimedia learning are scarce. In this study, due to the small sample sizes, only White students who made up 60% of the participants were compared with other ethnicities. This study contributes to the few existing studies such as Moreno and Flowerday (2006) that have examined ethnic differences in multimedia learning.

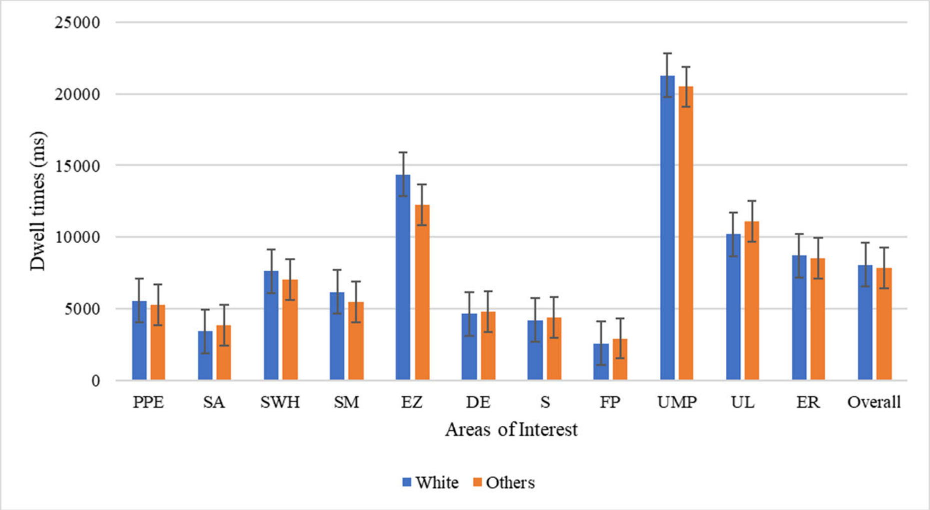


Fig. 8: Dwell Time for Annotated Video Based on Ethnicity

## 4.3 Cognitive Load

As shown in Figure 9, the participants reported higher effort and frustration for the annotated video, however, they reported lower mental demand for the same. This shows that it was not mentally demanding to learn with the annotated video, but participants put in more effort and experienced higher frustration. This could be because of their unfamiliarity with the annotated video, the effect could be abated as learners get used to the video. Overall, the participants experienced a higher cognitive load in the annotated learning environment. No significant difference was observed between the annotated and unannotated video ($p > 0.05$). This result differs from Chiu et al. (2018) who reported that students who learned with annotated video experienced lower cognitive load. This difference could be attributed to other factors such as participants' ethnicity (Moreno & Flowerday, 2006), visual preference (Homer et al., 2008) and media type (Castro-Alonso et al., 2019) which are moderating variables in multimedia learning. For instance, Homer et al. (2008) reported that low visual-preference learners experienced higher cognitive load than high visual-preference learners in learning with video.

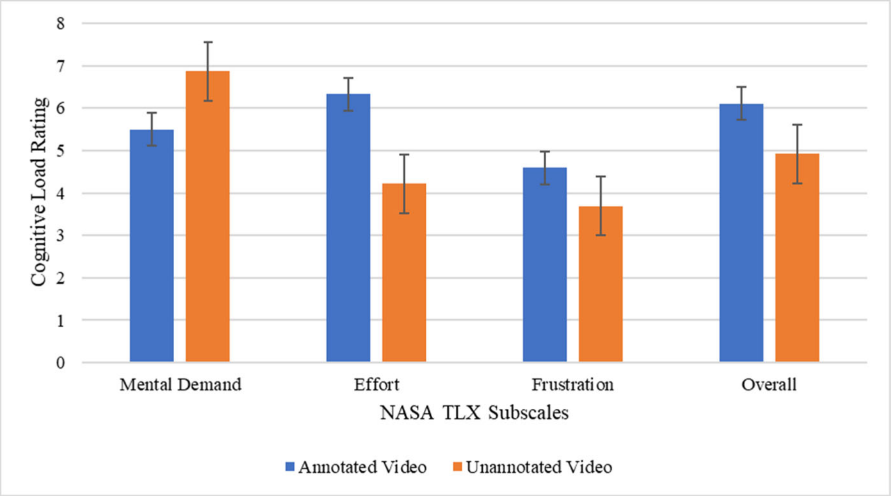


Fig. 9: Participants' ratings of perceived mental demand, effort, and frustration after interaction with both video

The demographic comparison of cognitive load in annotated video is shown in Figure 10. Although no statistically significant difference in any of the comparisons ($p$>0.05), female participants reported higher cognitive load compared to male participants. Similarly, participants with 0-2 years of experience self-reported higher cognitive load than their colleagues with 2-5 years of experience. For the academic programs, CEE students reported the highest cognitive load, followed by CEM students, while BC students had the lowest cognitive load rating. Demographic comparison reveals variations in the cognitive load level of the students. This variation could help instructors to design and adapt multimedia learning environments to suit learners of various categories. This is especially important to instructors especially those in engineering-construction educational programs, where effort is required to attract and retain female high school students and those from underrepresented groups (Choi et al., 2022).

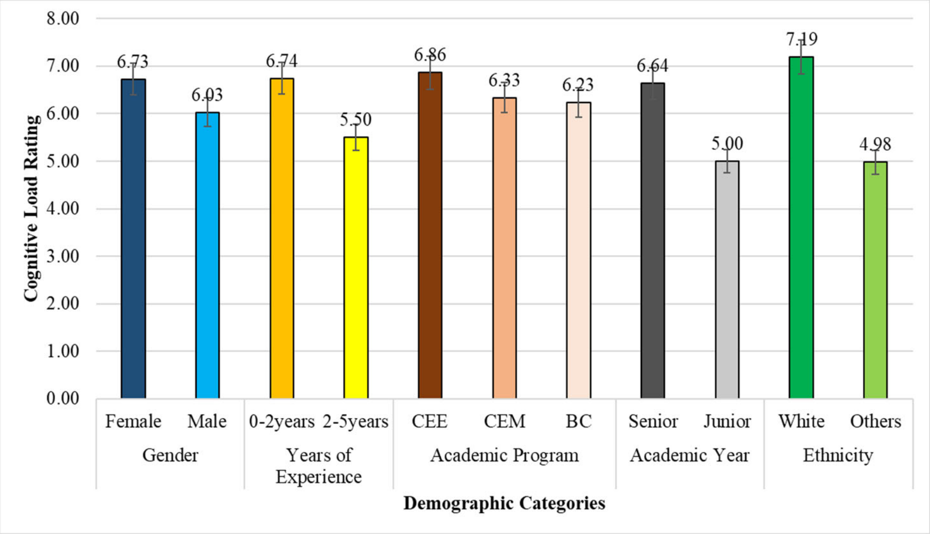


Fig. 10: Demographic comparison of cognitive load in annotated video

## 4.4 Verbal Feedback

Most of the participants that watched the annotated video opined that the annotation was effective in making them learn easier. The participants reported that the annotation helped them to focus on learning content and their attention was held. This is because in addition to the audio narration, the annotation helped the student to easily identify specific areas to focus on in the video which helped them to better understand the learning contents. Only three participants reported that the signals were distracting. The participants suggested using arrows instead of boundary boxes and highlighting their content because some learners might focus on the boundary boxes more than the contents within them. Some participants also suggested making the video more interactive and reducing the tempo. The students attested to the potential of the video for online courses and helping to learn about more construction practices without having to visit a job site.

## 5. CONCLUSION, LIMITATIONS AND FUTURE WORK

Characteristics of learners influence cognitive load and the efficacy of designs in multimedia learning. Hence, it is important to evaluate multimedia pedagogical tools to assess their suitability for intended context, purpose, and users. This study evaluates the efficacy of an annotated video which is a computer-vision-aided multimedia learning for teaching construction safety practices. The study reveals that computer vision generated signals were effective in drawing the intention of learners to fixate on AOIs. Within the annotated learning environment, female students, students with 0-2 years of experience, senior-level students, CEE students, and white students dwell longer and on more AOIs than their counterparts. However, these categories of participants reported higher overall cognitive load for the annotated video compared to their counterparts. The study also shows demographic differences in the cognitive load level of participants based on gender, ethnicity, academic level, years of experience and academic program. The results reveal that the demographic classes that dwelled more on the AOIs also reported a higher cognitive load. The results of this study could help instructors in engineering-construction education programs to effectively use annotated videos as instructional materials. This study could serve as a benchmark for future studies on artificial intelligence techniques for signaling in multimedia learning. The study opened a discussion on demographic differences in multimedia learning within the construction engineering education domain as well as the efficacy of artificial intelligence techniques in the design of multimedia pedagogical tools. This study has some limitations which could be the focus of future research. For example, only subjective rating of cognitive load was used, future research could combine both objective and subjective measures to assess cognitive load level in multimedia learning. Also, the small sample size of this study could have been responsible for some lack of significant differences in the comparisons made across the demographics of the participants. The small sample size also limits the generalizability of the findings. Future research could use higher sample sizes. Also, effects of age differences and academic levels (i.e., elementary school, high school, and college) of participants in multimedia learning could be the subject of future work.

## REFERENCES

Abdulrahaman, M., Faruk, N., Oloyede, A., Surajudeen-Bakinde, N., Olawoyin, L., Mejabi, O., Imam-Fulani, Y., Fahm, A., & Azeez, A. (2020). Multimedia tools in the teaching and learning processes: A systematic review. *Heliyon*, *6*(11).

Abudayyeh, O., Russell, J., Johnston, D., & Rowings, J. (2000). Construction engineering and management undergraduate education. *Journal of construction engineering and management*, *126*(3), 169-175. https://doi.org/https://doi.org/10.1061/(ASCE)0733-9364(2000)126:3(169).

Afonso, A., Ramírez, J. J., & Díaz-Puente, J. M. (2012). University-industry cooperation in the education domain to foster competitiveness and employment. *Procedia-Social and Behavioral Sciences*, *46*, 3947-3953. https://doi.org/https://doi.org/10.1016/j.sbspro.2012.06.177.

Aronson, J. D. (2018). Computer vision and machine learning for human rights video analysis: Case studies, possibilities, concerns, and limitations. *Law & Social Inquiry*, *43*(4), 1188-1209.

Bojko, A. (2013). *Eye tracking the user experience: A practical guide to research*. Rosenfeld Media.

Bozoglu, J. (2016). Collaboration and coordination learning modules for BIM education. *J. Inf. Technol. Constr.*, *21*, 152-163.

Carter, B. T., & Luke, S. G. (2020). Best practices in eye tracking research. *International Journal of Psychophysiology*, *155*, 49-62.

Castro-Alonso, J. C., Wong, M., Adesope, O. O., Ayres, P., & Paas, F. (2019). Gender imbalance in instructional dynamic versus static visualizations: A meta-analysis. *Educational Psychology Review*, *31*, 361-387.

Chiu, P.-S., Chen, H.-C., Huang, Y.-M., Liu, C.-J., Liu, M.-C., & Shen, M.-H. (2018). A video annotation learning approach to improve the effects of video learning. *Innovations in Education and Teaching International*, *55*(4), 459-469.

Choi, J. O., Shane, J. S., & Chih, Y.-Y. (2022). Diversity and inclusion in the engineering-construction industry. *Journal of Management in Engineering*, *38*(2).

De Koning, B. B., Tabbers, H. K., Rikers, R. M., & Paas, F. (2009). Towards a framework for attention cueing in instructional animations: Guidelines for research and design. *Educational Psychology Review*, *21*, 113-140.

Dousay, T. A., & Trujillo, N. P. (2019). An examination of gender and situational interest in multimedia learning environments. *British Journal of Educational Technology*, *50*(2), 876-887.

Eiris Pereira, R., & Gheisari, M. (2019). Site visit application in construction education: A descriptive study of faculty members. *International Journal of Construction Education and Research*, *15*(2), 83-99. https://doi.org/https://doi.org/10.1080/15578771.2017.1375050

Ferreira, C., Baptista, M., & Arroio, A. (2013). Teachers' pedagogical strategies for integrating multimedia tools in science teaching. *Journal of Baltic Science Education*, *12*(4), 509.

Grimley, M. (2007). Learning from multimedia materials: The relative impact of individual differences. *Educational Psychology*, *27*(4), 465-485.

Gunhan, S. (2015). Collaborative learning experience in a construction project site trip. *Journal of Professional Issues in Engineering Education and Practice*, *141*(1), 04014006. https://doi.org/https://doi.org/10.1061/(ASCE)EI.1943-5541.0000207.

Hashimoto, D. A., Rosman, G., Witkowski, E. R., Stafford, C., Navarrete-Welton, A. J., Rattner, D. W., Lillemoe, K. D., Rus, D. L., & Meireles, O. R. (2019). Computer vision analysis of intraoperative video: automated recognition of operative steps in laparoscopic sleeve gastrectomy. *Annals of surgery*, *270*(3), 414.

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.

Kalyuga, S. (2013). Effects of learner prior knowledge and working memory limitations on multimedia learning. *Procedia-Social and Behavioral Sciences*, *83*, 25-29.

Kalyuga, S., Chandler, P., & Sweller, J. (2000). Incorporating learner experience into the design of multimedia instruction. *Journal of Educational Psychology*, *92*(1), 126.

Kelley, K., Clark, B., Brown, V., & Sitzia, J. (2003). Good practice in the conduct and reporting of survey research. *International Journal for Quality in health care*, *15*(3), 261-266.

Latifzadeh, K., Amiri, S., Bosaghzadeh, A., Rahimi, M., & Ebrahimpour, R. (2020). Evaluating cognitive load of multimedia learning by eye-tracking data analysis. *Technology of Education Journal (TEJ)*, *15*(1), 33-50.

Law, E. L.-C., Mattheiss, E. E., Kickmeier-Rust, M. D., & Albert, D. (2010, November 4-5, 2010.). Vicarious learning with a digital educational game: Eye-tracking and survey-based evaluation approaches. 6th Symposium of the Workgroup Human-Computer Interaction and Usability Engineering, USAB 2010., Klagenfurt, Austria.

Mayer, R. E., & Fiorella, L. (2014). 12 principles for reducing extraneous processing in multimedia learning: Coherence, signaling, redundancy, spatial contiguity, and temporal contiguity principles. In *The Cambridge handbook of multimedia learning* (Vol. 279). Cambridge University Press New York, NY.

Molina, A. I., Navarro, Ó., Ortega, M., & Lacruz, M. (2018). Evaluating multimedia learning materials in primary education using eye tracking. *Computer Standards & Interfaces*, *59*, 45-60.

Moreno, R., & Flowerday, T. (2006). Students' choice of animated pedagogical agents in science learning: A test of the similarity-attraction hypothesis on gender and ethnicity. *Contemporary educational psychology*, *31*(2), 186-207.

Navarro, O., Molina, A. I., Lacruz, M., & Ortega, M. (2015). Evaluation of multimedia educational materials using eye tracking. *Procedia-Social and Behavioral Sciences*, *197*, 2236-2243.

Olayiwola, J., Akanmu, A., Gao, X., Murzi, H., & Afsari, K. (2023). Design and Usability Evaluation of an Annotated Video–Based Learning Environment for Construction Engineering Education. *Journal of Computing in Civil Engineering*, *37*(6), 04023033.

Olayiwola, J., Yusuf, A. O., Akanmu, A. A., Murzi, H., Gao, X., & Afsari, K. (2023). Construction practice knowledge for complementing classroom teaching during site visits. *Smart and Sustainable Built Environment*, *Ahead-of-print*(Ahead-of-print), Ahead-of-print.

Pedro, A., Le, Q. T., & Park, C. S. (2016). Framework for integrating safety into construction methods education through interactive virtual reality. *Journal of Professional Issues in Engineering Education and Practice*, *142*(2), 04015011.

Poonja, H. A., Shirazi, M. A., Khan, M. J., & Javed, K. (2023). Engagement detection and enhancement for STEM education through computer vision, augmented reality, and haptics. *Image and Vision Computing*, *142*(2), 104720.

Refat, N., Kassim, H., & Rahman, M. A. (2020). A cognitive approach-based instructional design for managing cognitive load and improving learning outcome. 2020 Emerging Technology in Computing, Communication and Electronics (ETCCE), Bangladesh.

Saha, S., & Halder, S. (2016). He or She: Does gender affect various modes of instructional visual design? *Journal of Research on Women and Gender*, *7*(1), 47-58.

Savov, T., Terzieva, V., & Todorova, K. (2018). Computer vision and internet of things: Attention system in educational context. Proceedings of the 19th International Conference on Computer Systems and Technologies,

Sharek, D. (2011). A useable, online NASA-TLX tool. Proceedings of the human factors and ergonomics society annual meeting,

Shojaei, A., Rokooei, S., Mahdavian, A., Carson, L., & Ford, G. (2021). Using immersive video technology for construction management content delivery: a pilot study. *J. Inf. Technol. Constr.*, *26*, 886-901.

Sophokleous, A., Christodoulou, P., Doitsidis, L., & Chatzichristofis, S. A. (2021). Computer vision meets educational robotics. *Electronics*, *10*(6), 730.

Stark, L., Brünken, R., & Park, B. (2018). Emotional text design in multimedia learning: A mixed-methods study using eye tracking. *Computers & Education*, *120*, 185-196.

Tang, S., Roberts, D., & Golparvar-Fard, M. (2020). Human-object interaction recognition for automatic construction site safety inspection. *Automation in Construction*, *120*, 103356.

Tetiana, M., Kondratenko, Y., Sidenko, I., & Kondratenko, G. (2021). Computer vision mobile system for education using augmented reality technology. *Journal of Mobile Multimedia*, *17*(4), 555–576.

Tudić, V., Stančić, A., Kralj, D., & Tropčić, T. (2022). Application of Computer Vision in Education in Mechatronic Control System. 2022 ELEKTRO (ELEKTRO), Krakow, Poland.

Zhang, H., Kyaw, Z., Chang, S.-F., & Chua, T.-S. (2017). Visual translation embedding network for visual relation detection. Proceedings of the IEEE conference on computer vision and pattern recognition.

Zhao, Q., & Li, Z. (2021). Application of computer vision media simulation technology in distance education of new generation labor productivity. Journal of Physics: Conference Series.