

Effectiveness of Video-based Training for Face-to-Face Communication Skills of Software Engineers: Evidence from a Three-Year Study

Antonija Mitrovic*

University of Canterbury, Christchurch, New Zealand, tanja.mitrovic@canterbury.ac.nz

Matthias Galster

University of Canterbury, Christchurch, New Zealand, matthias.galster@canterbury.ac.nz

Sanna Malinen

University of Canterbury, Christchurch, New Zealand, sanna.malinen@canterbury.ac.nz

Jay Holland

University of Canterbury, Christchurch, New Zealand, jay.holland@canterbury.ac.nz

Ja'afaru Musa

University of Canterbury, Christchurch, New Zealand, jaafarum1@gmail.com

Negar Mohammadhassan

University of Canterbury, Christchurch, New Zealand, negar.mohammadhassan@pg.canterbury.ac.nz

Raul Vincent Lumapas

University of Canterbury, Christchurch, New Zealand, raulvincent.lumapas@pg.canterbury.ac.nz

Objectives. Communication skills are crucial for effective software development teams, but those skills are difficult to teach. The goal of our project is to evaluate the effectiveness of teaching face-to-face communication skills using AVW-Space, a platform for video-based learning which provides personalized nudges to support student's engagement during video watching.

Participants. The participants in our study are second-year software engineering students. The study was conducted over three years, with students enrolled in a semester-long project course.

Study Method. We performed a quasi-experimental study over three years to teach face-to-face communication using AVW-Space, a video-based learning platform. We present the instance of AVW-Space we developed to teach face-to-face

communication. Participants watched and commented on ten videos, and later commented on the recording of their own team meeting. In 2020, the participants (n = 50) did not receive nudges, and we use the data collected that year as control. In 2021 (n = 49) and 2022 (n = 48), nudges were provided adaptively, to encourage students to write more and higher-quality comments.

Findings. The findings from the study show the effectiveness of nudges. We found significant differences in engagement when nudges were provided. Furthermore, there is a causal effect of nudges on the interaction time, the total number of comments written and the number of high-quality comments, as well as on learning. Finally, participants exposed to nudges reported higher perceived learning.

Conclusions. Our research shows the effect of nudges on student engagement and learning while using the instance of AVW-Space for teaching face-to-face communication skills. Future work will explore other soft skills, as well as providing explanations for the decisions made by AVW-Space.

CCS CONCEPTS • **Applied computing** → **Interactive learning environments**; • **Information systems** → **Personalization**

Additional Keywords and Phrases: Teaching soft skills, video-based learning, AVW-Space, nudges, causal effects, student perceptions

ACM Reference Format:

Antonija Mitrovic, Matthias Galster, Sanna Malinen, Jay Holland, Ja'afaru Musa, Negar Mohammadhassan and Raul Vincent Lumapas. 2023. Effectiveness of video-based training for face-to-face communication skills: Evidence from a three-year study. ACM Transactions on Computing Education. ACM, New York, NY, USA.

1 INTRODUCTION

1.1 Problem

Software engineers need not only technical skills (e.g., about programming and technologies), but also soft skills which enable effective team work and communication with different stakeholders in a project. For example, communication skills are crucial in software engineering (SE) to elicit and share information with various stakeholders such as other developers and end users [53]. Face-to-face meetings are especially important as they allow team members to share information and make decisions [19]. Accreditation bodies and universities acknowledge the importance of soft skills [26,51]. However, teaching these skills is time-consuming, requires hands-on exercises and are therefore not easily scalable [3,25]. SE education predominantly fosters soft skills training in group projects [58] since exercising soft skills needs real project work with diverse teams and constant feedback and guidance from instructors.

1.2 Review of Relevant Scholarship

1.2.1 Teaching Soft Skills in Software Engineering Education and Beyond

Soft skills can be practiced in classrooms using debates, role playing and demonstrations, as well as in case studies, field visits and mock interviews [64]. One common approach for teaching soft skills in software engineering education is to develop dedicated courses [2,49], although students do not always appreciate such courses [57]. Similarly, frequently used approach in software engineering education is project-based learning (PBL) [12], where

students work as a team to build a software system [34] in an educational setting. PBL allows students to practice soft skills [51,40]. Recent work has explored how soft skills such as diversity impact student performance in team projects [29].

Another approach to teaching soft skills is having students record themselves while using particular soft skills. For example, outside software engineering, Rehear et al. [55] used role-play videos to provide feedback to first-year dentistry students on their communication skills and how to improve them. Although such approaches can be effective, they take significant resources in terms of teaching time. Software engineering and computing courses often contain assessment items that require soft skills, but there is no explicit teaching of such skills due to already full curricula.

There have also been approaches to teaching soft skills using game-based learning, both in physical and virtual spaces. Game-based learning (GBL) uses various gaming technologies to create entertaining environments to promote learning [62]. GBL approaches for teaching soft skills range from using physical environments such as board games, e.g. [59], and educational escape rooms [4], to digitized environments. For example, Clark and colleagues [13] describe educational escape rooms in which students work in teams to solve a problem (disarming a bomb) by solving multiple tests required to release an engineer hostage, with the main goal being improving communication, leadership and teamwork skills. Digitalized environments include simulations which allow students to practice certain task while at the same time practicing their soft skills [28]. The popularity of computer games gave rise to *serious games*, i.e. games primarily designed for learning rather than entertainment [22]. McGowan and colleagues [42] describe *Compete!*, a single player 2D serious game that requires final-year university students to use several soft skills (creative problem solving, communication, stress management and teamwork) while working on sustainability problems. The authors also point out the risks involved with serious games, as they are often oversimplifications of real-life situations thus making soft skills difficult to translate to real world. Furthermore, the lack of engagement can lead to shallow learning.

A way to overcome challenges related to teaching soft skills is adopting Video-Based Learning (VBL). Previous studies show that VBL is appropriate for teaching soft skills [3,15,17,20]. Watching videos enables learners to recall their past experiences, and also to see other people using soft skills in various situations, thus facilitating reflection. We discuss the use of VBL for teaching soft skills further in Section 1.2.3.

1.2.2 Engagement

It is widely acknowledged that passive learning is inferior to learning actively. In educational research, active learning is defined as a set of activities which the student performs on top of passive listening [5]. Active learning is student-centered, as opposed to teacher-centered classroom education. In active learning, the role of the teacher becomes that of a facilitator, rather than the provider of information. Various types of activities can be used to engage students actively: asking questions, solving problems, debating with peers, think-pair-share exercises, case studies, peer teaching, peer review and others. In active learning, students perform activities and at the same time reflect on their knowledge, learn from each other and improve their critical thinking skills, group work and other soft skills.

Our research is based on the ICAP theory of cognitive engagement [9,10], which describes engagement based on the overt behaviors students enact while learning. ICAP identifies four categories of overt behaviors with decreasing level of engagement: *Interactive (I)*, *Constructive (C)*, *Active (A)* and *Passive (P)*. The theory states that the more engaged students are, the more they learn ($I > C > A > P$). Each learning mode has its own characteristic

behaviors. Examples of the passive mode include listening to a lecture, watching a video or reading a text, without performing any other activity. Such behaviors generally result in storing the newly received information, the outcomes of which would be that the student can recall that information. In the Active mode, students might be taking verbatim notes while being in a lecture or watching a video; such behaviors help integrate the new information with existing knowledge. The constructive mode goes further, as the student is inferring relationships with existing knowledge. An example of the constructive mode for a student who is attending a lecture and taking notes is when the student draws concept maps, self-explaining the material to themselves or drawing analogies, thus enabling transfer. In the constructive mode, the student is generating new information, or makes associations between the lecture/video content and their own experiences. The characteristic feature of the interactive mode is dialoguing: the student may ask for clarifications, comparing the new information, discussing material with peers or teachers, so that knowledge is co-created. For the interactive mode, all learners participating in the dialogue contribute to the knowledge creation. The interactive mode also includes specific types of interactions between learners and computer-based learning environment, assuming that there is turn taking and what is discussed is generating new information. Chi and Wiley [9,10] admit that overt behaviours are not a perfect indicator of how the student is learning; the student may be just listening to a lecture, without any overt behaviours, but internally could be self-explaining, making analogies or reflecting on their knowledge, which means that the student is engaging with the material in a constructive way. However, overt behaviours (or lack of them) are much easier to observe and serve as a good proxy for classifying learning situations. Chi and Wiley [9] also state that the boundaries between the four learning modes are blurry, and that the student may engage in multiple learning modes in one session.

The ICAP theory has been validated empirically in numerous studies [9,10]. The results of the prior studies with AVW-Space were in line with the ICAP theory [13,47], as discussed in the following section.

1.2.3 Supporting engagement in VBL

In online learning, videos are often the primary source of information, but require students to engage with presented material in order to learn. Mariachi et al. [44] note the similarity between watching videos and attending a lecture: in both cases the student needs to be attentive, think about concepts being discussed, and integrate the newly acquired knowledge with existing knowledge. Problems identified with VBL include limited interactivity with videos (including difficulties with navigating and searching), the lack of human interaction, personalization, assessment and feedback [8,14].

Just passively watching videos is not enough for learning [35,44]. Deep engagement with videos can be supported in several ways, starting with the careful design of videos themselves [6,30] to lower the cognitive load during VBL, by using segmenting and signaling to highlight important parts of the video. Significant work has been done on increasing engagement with videos, by adding annotation tools, quizzes, examples and interactive exercises [35,36,52,63]. For example, Kovacs [36] found that students engage heavily with quizzes, as they see the benefit in testing themselves and receiving immediate feedback. The author also found that in-video quizzes can influence students' video-viewing behavior, with some students focusing predominantly on the quizzes and seeking answers to the questions in preceding sections. Mariachi et al. [44] compared a group of learners who passively watched videos to two groups using active learning strategies: video annotations and in-video prompts. They found that students who annotated videos had higher self-efficacy and higher learning. These kinds of approaches result in improved engagement and learning, but require substantial effort by teachers.

Data-driven approaches using interaction traces from VBL have been proposed to improve techniques for video navigation, such as visualizations of collective navigation traces, dynamic timeline scrubbing and enhanced in-video searches [8]. There are also approaches using students' ratings, annotations and contributions to forums to support social navigation and collaborative learning [8,54]. Hyper video players provide advanced controls for navigation through multiple videos, links to supplementary material and support for collaboration. For example, Dodson et al. [23] present a study with the Vide hypervideo player, in which students could watch lecture recordings. Vide provides visual representation of the video content in the form of a filmstrip, as well as the video transcript. The student can take notes and highlight parts of the video transcript. Vide also provides histograms below videos, displaying how much time the student spent on various parts of the video. The student can navigate through the video by using the filmstrip or transcript. Note Struct [41] prompts users to take notes, and later engages learners in reflecting and summarizing their notes. Durrell [14] extracts concepts from videos and organizes them in knowledge graphs, thus allowing students to explore the video content more effectively. Other hypermedia players provide advanced control for navigation through multiple videos, link to supplementary material and support for collaboration [7].

Active Video Watching (AVW) was suggested as a VBL approach that encourages reflective learning [16,21,39,45,61,]. The AVW-Space platform implements AVW by allowing instructors to embed YouTube videos for students to watch and comment on. The platform supports reflection during interactive note taking by requiring students to specify one of the teacher-specified aspects to each comment [21]. The aspects focus students' attention on their previous experience or thinking about the future performance. The platform also supports social learning through rating of comments written by other students in the class.

Early studies with AVW-Space show that students who write comments and rate comments written by other students increase their knowledge. On the other hand, there was a substantial group of students who do not engage with videos actively [45]. Therefore, nudges were added to AVW-Space in order to increase engagement [20,21]. Nudges are used to encourage students to write comments, while at the same time preserving the student's freedom to ignore nudges and use videos in their preferred way.

There are two types of nudges in AVW-Space: reminder nudges and quality nudges. The goal of reminder nudges is to encourage students to write comments, and such nudges are given to students who passively watch videos [20,46]. The other type of nudges are quality nudges. AVW-Space automatically classifies comments as students write them into three categories: low quality (brief, vague or off-topic comments), medium quality (comments that repeat the video content) and high-quality comments (i.e. comments with elaboration, reflection or planning for the future) [47]. Quality nudges are given when students write low-quality comments, and provide examples of high-quality comments or encourage students to write about a particular topic. These nudges also encourage students who have written high-quality nudges to continue writing them. Visualizations were also added to AVW-Space, to provide information about the student's performance on different activities to help them reflect on their learning [48]. The results of the studies show that the percentage of constructive students was higher when nudges and visualizations were provided to students.

While video-based learning has been used to teach programming skills (e.g., by providing tutorial videos to students or recording videos for later viewing), video-based learning of soft skills is not common in software engineering.

1.3 Aims and Research Questions

Prior research on AVW-Space focused on presentation skills as the target soft skill [21,47]. The research reported in this paper investigates the use of AVW-Space to teach a new soft skill: communication in face-to-face software development meetings. Face-to-face meeting communication skills are particularly relevant for software engineering practitioners [27] since practitioners spend a significant amount of their time meeting and working with others (rather than sitting down to write code on their own). For example, software engineers spend time in planning meetings, working meetings with the whole team or parts of the team, review meetings, or one-on-one's with team mates, customers or managers (and collaboration, psychological safety, transparency and accountability strengthen the team's ability to produce quality software) [1]. Please note that AVW-Space is freely available to teachers wanting to develop spaces for their classes (<https://www.canterbury.ac.nz/engineering/schools/csse/research/ictg/avw-space/>).

We developed a new instance of AVW-Space for this soft skill, and conducted a three-year long study in the second-year software engineering project course at the University of Canterbury, New Zealand. Based on a quasi-experimental design, we used the year 2020 as control, with AVW-Space providing just support for commenting on videos and rating comments written by peers. In 2021 and 2022, students had full support, with AVW-Space providing personalized nudges and visualizations based on students' behavior. In addition to providing YouTube videos on the target skill, we also recorded team meetings and provided the recordings in AVW-Space for students to review. Our study addresses the following research questions:

RQ1: *What are the differences in student engagement and learning in the instances of the course with or without nudges?*

Since previous research with AVW-Space on presentation skills shows that nudges were effective in increasing students' engagement and learning, we were interested in evaluating whether similar effects would be found when teaching a different soft skill, face-to-face communication skills.

RQ2: *Is there a causal effect of nudges on students' engagement and learning?*

We wanted to determine explicitly the effect of nudges on engagement and learning, by conducting causal modeling.

RQ3: *What kind of perceptions do students have on usefulness of the recordings of team meetings?*

Since students work in teams in the chosen SE course, we wanted to determine whether students would be interested in watching their own team meeting, as well as finding out what kind of reflections and learning result from such activity.

RQ4: *What are students' perceptions of learning in AVW-Space?*

We were interested in finding the students' perception of the provided online training, as well as to investigate whether there are any differences in students' perceptions when nudges were or were not provided.

2 METHOD

We implemented a new instance of AVW-Space for teaching face-to-face communication skills in software development meetings, and integrated this online training into a second-year semester-long SE project course at the University of Canterbury in 2020-2022. We obtained approval for the research from our institution's Human Ethics committee. In the course, the students worked in teams of four to six, had weekly face-to-face meetings, and received a short lecture on what to do before, during and after each meeting. In addition, the students were invited

to use AVW-Space to learn face-to-face meeting communication skills. Students received 5% of the final grade if they completed all phases of the online training.

Table 1: Overview of the study

Week	Tasks
1	Forming groups
2	Survey 1; watch and comment on videos
3	Review and rate comments written by other students in the class
4	Meeting recording (Survey 2 in 2021/2022)
5	Watch and comment on the team meeting
6	Review and rate comments on the team meeting
7	Survey 3

The study was done over the initial seven weeks of the course, as specified in Table 1. In week 1, the lecturers formed the groups and informed students about the online training. Students completed Survey 1 and then watched and commented on the videos in week 2. We selected ten publicly available videos from YouTube (listed in Table 2), and integrated them into AVW-Space. In all three years, the same ten videos were provided to the students. Six tutorial videos cover various elements of face-to-face communication, while four example videos present real (or acted) meetings. The students were instructed to watch and comment on tutorial videos first, and later to use what they have learnt to critique the example videos. The screenshot in Figure 1 shows a student writing a comment on a tutorial video. When writing a comment, the video is paused. The student can enter the text of the comment in the provided box, and also needs to select one of the aspects (listed above the comment box), to specify what the comment is related to. Please note that we used the same aspects for writing comments based on tutorial videos as in the previous AVW-Space studies done on presentation skills [21]. Three aspects were defined to stimulate learners to reflect on their experience: *“I am rather good at this,” “I didn’t realize I wasn’t doing it,” “I did/saw this in the past”* and the final aspect was *“I like this point.”* We defined the following aspects for writing comments on example videos: *“verbal communication”, “giving feedback”, “receiving feedback”, “active listening”* and *“meeting contributions”*, corresponding to the concepts introduced in the tutorial videos. The screenshot in Figure 1 also shows a nudge (the red box titled *“Make a comment”*). This nudge was given as the student has watched the video before but has not written any comments.

Table 2: The videos used in the study

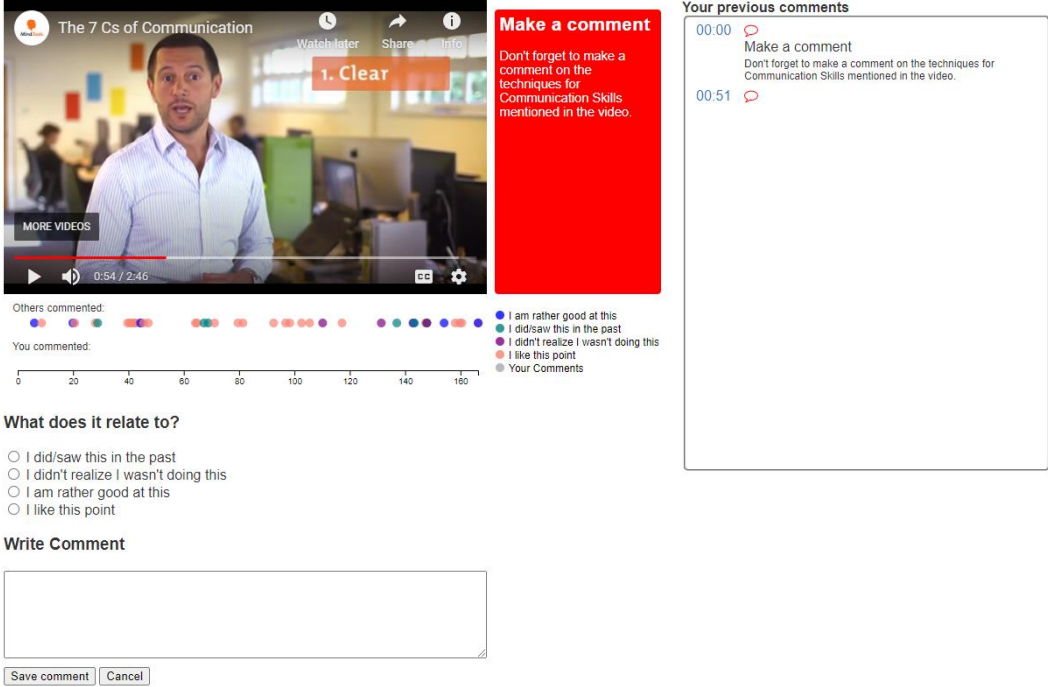
Video	Title	Length	YouTube video id
Tutorial 1	The 7 Cs of Communication	2:46'	sYBw9-8eCuM
Tutorial 2	Body Language	2.45'	AqixzdpJL4U
Tutorial 3	Giving feedback	1.46'	Id_uG8Djdsc
Tutorial 4	Improve your listening skills with active listening	2.39'	t2z9mdX1j4A
Tutorial 5	How Google builds the perfect team	2.22'	v2PaZ8NI2T4
Tutorial 6	How to effectively contribute to team meetings	4.05'	cKh75Po5Qsc
Example 1	Bad Stand-up	5.22'	zrmcl-pjmoc
Example 2	The Daily Stand-up Meeting	2.34'	VjNxQ-a-x2M
Example 3	Examples of Good Meeting Communications Skills	1.50'	czpBKC9Plh4
Example 4	How NOT to run a meeting	2.37'	F1qstYxrqn8

In week 3 anonymized comments were made available to the whole class for rating. Students could rate comments using the rating categories previously used in [21]: *“This is useful for me”, “I hadn’t thought of this”, “I didn’t notice this”, “I don’t agree with this”* and *“I like this point”*.

Following that, in week 4 students received Survey 2 in 2021 and 2022 (please note that in 2020 there was no such survey). Additionally, one meeting of each group was recorded and added to AVW-Space. Only members of

the team could watch their own team meeting. Students were asked to watch and comment on their own meeting in week 5, and then rate each others' comments in week 6. In week 7, students were asked to complete the final survey (referred to as Survey 3).

Figure 1. Comment writing in AVW-Space



2.1 Participant Characteristics

Table 3 presents demographic data about our participants. The class size is shown in the *Course* column, while the next column shows the number of students who completed Survey 1 (2020: 50 students, 2021: 49 students, 2022: 48 students). There were no students who repeated the course. In all three years, most of the participants were male, in the 18-23 age group and were native English speakers, which is typical for SE classes at our institution. Most participants had no formal training on communication in face-to-face meetings. Those participants who had formal training got it in high school or University, or received training at community/volunteer group. There were also three participants who received professional development training. Most participants had no software engineering experience outside their University courses, but all passed introductory programming courses and an introductory course to software engineering (including software engineering principles, practices, processes and tools, and object-oriented programming and design, as well as a small software project completed in pairs) There were also two Likert questions on how often the participants watched YouTube (the *YouTube* column), and how often they watched YouTube videos for learning (the *YT4L* column). For these two questions, the options were 0 (never), 1 (occasionally), 2 (once a month), 3 (every week) or 4 (every day). The averages and standard deviations

presented in Table 3 show the high usage of YouTube; none of the participants responded with 0 for those questions.

Table 3: Demographic data

Year	Course	n	Male	18-23 age	Native English	No training	No SE experience	YouTube mean (SD)	YT4L mean (SD)
2020	56	50	84%	98%	78%	80%	74%	3.38 (.95)	2.26 (.99)
2021	50	49	80%	98%	88%	70%	80%	3.53 (.82)	2.73 (1.06)
2022	48	48	67%	92%	81%	75%	81%	3.63 (.64)	2.73 (.82)

2.2 Instrumentation and Data Collection

AVW-Space automatically logs all actions students perform while interacting with the platform. In addition to the logs, we collected data via three surveys (administered within AVW-Space). The surveys are provided in the Appendix.

Survey 1 was identical across the years, and contained demographic questions, and questions on training and experiences with face-to-face meetings. In addition, there was a timed question asking students to write everything they know about face-to-face meetings communication skills in three minutes. Surveys 2 and 3 also contained the same question about students' knowledge of face-to-face communication in meetings.

Survey 3 included three instruments: (1) CAP perceived learning scale [56], (2) NASA-TLX cognitive load scale [31], and (3) Technology Acceptance Model (TAM) scale [18] to capture students' overall perception of AVW-Space. Additionally, Survey 3 contained open-ended questions on students' perceptions of various features of AVW-Space.

The CAP perception of learning scale [56] was used to capture students' estimates of learning from their experience with AVW-Space. The scale consists of nine items organized into three dimensions: Cognitive, Affective and Psychomotor. These items were presented to students in Survey 3. We slightly modified the text of each item to suit our experimental design, as specified in Table 4. We did not use the second question as our study was much shorter than a full course. The students responded to items using a Likert scale from 0 (*strongly disagree*) to 6 (*strongly agree*). Please note that item 7 is negatively worded and hence we have reversed the score for that item.

The NASA-TLX task load index is a multidimensional scale assessing participants' workload [31]. The scale consists of six dimensions: mental, physical and temporal demands, frustration, effort and performance. Due to the nature of the tasks in our study, we have not used the physical and temporal dimensions.

The Technology Acceptance Model (TAM) is a validated instrument which measures perceived usefulness and perceived ease of use of a particular computer system [18]. Perceived usefulness is defined as "the degree to which a person believes that using a particular system would enhance his or her job performance", while the perceived ease of use is defined as "the degree to which a person believes that using a particular system would be free of effort" [18]. These two factors determine the overall acceptance of the system, and have been shown to be significantly correlated with both self-reported current or predicted future use of studied computer systems [18].

Table 4: The items of the CAP perceived learning scale

	Original Text	Dimension	Modified text
1	I can organize course material into a logical structure	Cognitive	I can summarize what I have learnt in AVW-Space for someone who has not learned from AVW-Space.
2	I cannot produce a course study guide for future students.	Cognitive	N/A
3	I am able to use physical skills learned in this course outside of class	Psychomotor	I am able to use the effective meeting participation concepts I learnt in AVW-Space in my future meetings.
4	I have changed my attitudes about the course subject matter as a result of this course.	Affective	I have changed my attitudes about effective meeting participation as a result of AVW-Space.
5	I can intelligently critique the texts used in this course.	Cognitive	I can assess the quality of face-to-face communication in the example videos used in this training.
6	I feel more self-reliant as the result of the content learned in this course.	Affective	I feel more confident in my face-to-face communication skills in meetings as a result of AVW-Space.
7	I have not expanded my physical skills as a result of this course.	Psychomotor	I have not expanded my knowledge of effective meeting participation concepts as a result of AVW-Space.
8	I can demonstrate to others the physical skills learned in this course	Psychomotor	I can demonstrate to others the effective meeting participation concepts I learnt in AVW-Space.
9	I feel that I am a more sophisticated thinker as a result of this course	Affective	I feel that I am a more effective meeting participant as a result of AVW-Space.

2.3 Measuring students' knowledge

The conceptual knowledge questions, which appeared in all three surveys, were used as a proxy for the test of the students' conceptual knowledge of the target skill. The instructions given to students were: "Write all words/phrases (one per line) that you associate with effective communication in software engineering meetings."

Students' responses were marked automatically, using text analytics methods and the vocabulary of the target skill developed by the authors. In order to develop the vocabulary, we first generated the corpus from the transcripts of tutorial videos. The tokens were extracted and lemmatized after lowercasing texts and removing punctuation and stop words. Next, using collocation statistics [43] implemented in the Phrases module of the Genism library 7, words and bigram phrases that appeared more than twice in the corpus were extracted automatically. In addition to collocation statistics, following the work of [50], we extracted the most relevant and similar words using Global Vectors (GloVe) Word Representation to represent each word. The total of 225 words and phrases were extracted along with 225 synonyms. Three independent expert coders verified whether the extracted words should be in the domain vocabulary. Each word was coded with 1 or 0, depending on whether a particular word was relevant or not. The pairwise Cohen's Kappa test revealed moderate (0.55), substantial (0.61) and nearly perfect (0.91) agreement between the coders [38]. Fleiss' kappa $\kappa = 0.69$ also showed substantial inter-coder agreement [38], but Krippendorff's alpha coefficient $\alpha = 0.31$ was low [37]. The three coders then reviewed their codes with a fourth coder to resolve differences, using the majority vote to achieve agreement (NB all coders are authors of this paper). As the result, eleven words were excluded, and ten new words were added. The conceptual knowledge score is the number of the vocabulary concepts that appear in a student's response.

2.4 Analytic Strategy

We used non-parametric statistical data analyses when data was not normally distributed. Repeated measures ANOVA was used to compare conceptual knowledge scores from the three surveys and across three years. Pairwise comparisons were performed with Bonferroni correction. All analyses were performed using IBM SPSS version 28.0.1.0 unless otherwise specified. For causal modeling, we used thinkCausal (apsta.shinyapps.io/thinkCausal/).

3 RESULTS

3.1 Analyzing students' engagement and learning

We removed the data about students who completed Survey 1 but had watched no videos from further analyses. To answer RQ1, we analyzed the collected data to generate measures of how students interacted with AVW-Space (Table 5). There are some measures which are statistically significantly different. For example, the number of sessions students had in AVW-Space and the number of distinct days on which they interacted with the platform are different in the three years, with 2021 and 2022 numbers being significantly higher in comparison to 2020. The *Videos* row in the table reports the average number of tutorial/example videos watched; please note that some students watched the same video multiple times, which explains why the reported numbers are higher than 10 (the total number of videos). The number of videos watched is significantly higher in 2021, although we do not have an explanation for such an increase.

More importantly, the number of comments written on the tutorial/example videos is significantly higher in 2021 and 2022 than in 2020, when there were no nudges and visualizations. There was no significant difference between the number of nudges students received in 2021 and 2022, which is to be expected as the experimental set up was identical in those two years.

The *High Quality (HQ) comments* column reports the number of high-quality comments written by students. Such comments show self-reflection, critical thinking about the video content, or planning for future performance. The categorization of comments is done automatically by AVW-Space immediately after a comment is written. AVW-Space uses the Machine Learning classifiers for categorizing comments as explained in [47]. We found a significant difference on the number of high-quality comments, with 2020 students writing significantly fewer HQ comments in comparison to 2021 and 2022. This finding is not surprising as nudges were designed in order to motivate students to write better quality comments [47].

Table 5: Interaction with AVW-Space: mean (SD)

	2020	2021	2022	Significance
Participants	47	49	47	
Sessions	5.55 (2.06)	6.67 (3.83)	9.00 (9.01)	F = 4.39, p < .05
Days	4.77 (1.67)	5.49 (2.60)	6.83 (3.05)	F = 8.19, p < .001
Videos	11.77 (3.54)	44.33 (18.08)	13.85 (3.95)	F = 132.35, p < .001
Comments	9.62 (13.21)	29.55 (17.81)	29.81 (18.54)	F = 22.76, p < .001
Nudges	N/A	44.33 (18.08)	45.94 (18.28)	
HQ comments	1.87 (2.58)	9.10 (6.30)	9.13 (6.25)	F = 28.89, p < .001
Ratings	104.47 (154.16)	145.24 (280.88)	58.77 (62.35)	no
CK1	6.77 (4.57)	7.15 (4.62)	8.77 (4.01)	no
CK2	N/A	8.59 (4.18), n = 44	11.71 (5.46), n = 41	F = 8.81, p < .01
CK3	9.90 (6.91), n = 30	10.70 (4.14), n = 10	10.20 (4.85), n = 40	no

The last three rows of Table 5 report the conceptual knowledge (CK) scores from Surveys 1, 2 and 3 (CK1, CK2 and CK3 respectively). There was no significant difference between the three classes on the initial (CK1) scores. In all three years, the final score (CK3) is significantly higher than CK1. In 2020, we did not have the Survey 2, and it was not possible to see whether the increase in the conceptual knowledge score is due to learning from the initial 10 videos, or due to the learning related to the recording of the team meeting. For that reason, we introduced Survey 2 in 2021 and 2022. Unfortunately, the COVID-19 lockdown overlapped with the last phase of the study in 2021, and the number of participants who completed that last phase was only 10. Therefore, we analyzed deeper the data from 2022. The repeated measures ANOVA for the conceptual knowledge scores for the 2022 data is significant ($F=4.53$, $p < .02$, partial eta squared = .104). The pairwise comparisons using the Bonferroni correction show a significant difference between CK1 and CK2 ($p = .03$), but no other significant differences. We found no significant differences between the CK3 scores of the three groups of participants.

Another way to analyze students' engagement is to classify them post-hoc based on their overt behaviors, using the ICAP framework [9]. As discussed previously, ICAP identifies four categories with decreasing level of engagement: *Interactive (I)*, *Constructive (C)*, *Active (A)* and *Passive (P)*. The ICAP framework states that the more engaged students are, the more they will learn ($I > C > A > P$). Interactive mode is irrelevant for our study as AVW-Space does not support direct interaction between students. Students classified as Passive are those who only watched videos and have not written any comments. We distinguish Constructive from Active students by observing the number of high-quality comments on tutorial videos, in line with [21,47]. The median number of high-quality comments on tutorial videos in 2020 was 1. Therefore, active students are those who wrote at most one high-quality comments, while constructive students wrote two or more such comments.

Table 6 presents the distribution of students over the three ICAP categories. In 2020, there were 16 students who only watched videos, without writing any comments, while in 2021 and 2022 all students wrote comments. Similar effect of nudges has been observed in previous studies on presentation skills in AVW-Space [21,47]. The number of Active students also reduced from 2020 to 2021/2022, when a vast majority of students were classified as Constructive. A chi-square test of homogeneity between the three years and ICAP categories revealed a significant difference (chi-square = 49.83, $p < .001$) with the effect size (Phi) of .59 ($p < .001$). A post-hoc analysis with the Bonferroni corrections revealed a significant increase of the number of Constructive students and a significant decrease of Passive students in 2021/2022 ($p < .0056$).

Table 6: The number of students in various ICAP categories

Year	Summary	n	Passive	Active	Constructive
2020	No nudges	47	16	9	22
2021	Nudges + Visualizations	49	0	3	46
2022	Nudges + Visualizations	47	0	1	46

To summarize our answer to RQ1, we found significant differences comparing student behavior in 2020 to 2021 and 2022, when nudges and visualizations were provided. In the latter case, a vast majority of students interacted more with AVW-Space and exhibited constructive behavior.

3.2 Causal Modeling

In order to identify causal effects, it is necessary to compare groups of participants who had or did not have a particular treatment, assuming random allocation of participants to groups. In real situations such as our study, it is not possible to control all factors that can potentially affect the outcome. Additionally, in educational studies it

is unethical to withhold treatment from some students. Instead, modeling the relationship between outcome variables, covariates and treatment is possible using the Bayesian Additive Regression Trees (BART) machine learning algorithm [11,32].

To determine the effect of nudges and visualizations on engagement and learning (RQ2), we conducted causal modeling using thinkCausal (apsta.shinyapps.io/thinkCausal/) [33], which uses BART as a foundation. This tool estimates the average treatment effects and is suitable for observational studies. Previous research shows that BART has excellent performance in causal inference [24,32]. The tool is user-friendly and scaffolds researchers through the data analytic process.

We grouped students from 2021 and 2022 together, as the experimental setup was identical in those two years. The results show that receiving nudges and visualizations led to an increase of 17.03 comments and an increase of 5.04 high-quality comments compared to what would have happened if students did not receive nudges (Figures 2 and 3 respectively). Similar results were obtained when we modeled other measures: There was an increase of 1.92 sessions when nudges/visualizations were provided. Providing visualizations and nudges also led to an increase of 0.52 conceptual knowledge scores CK3 in comparison to not providing them.

Figure 2. Causal modeling of the number of comments made

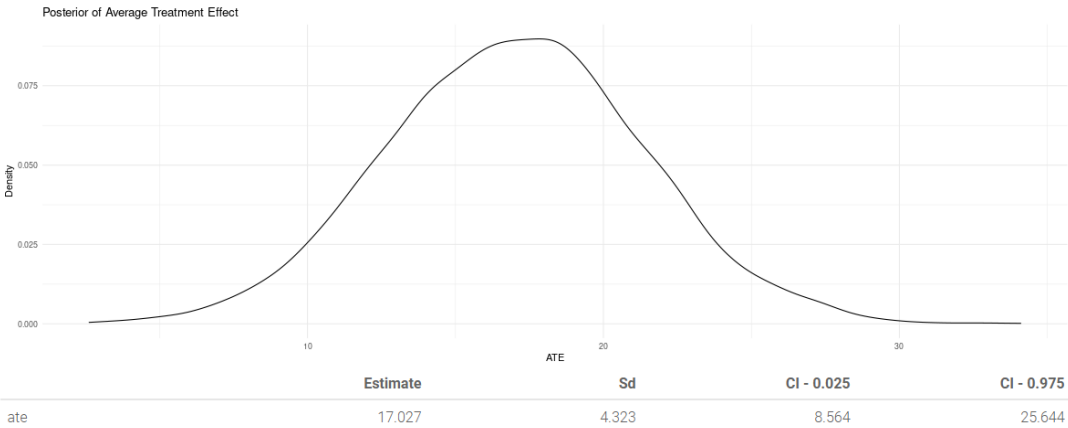
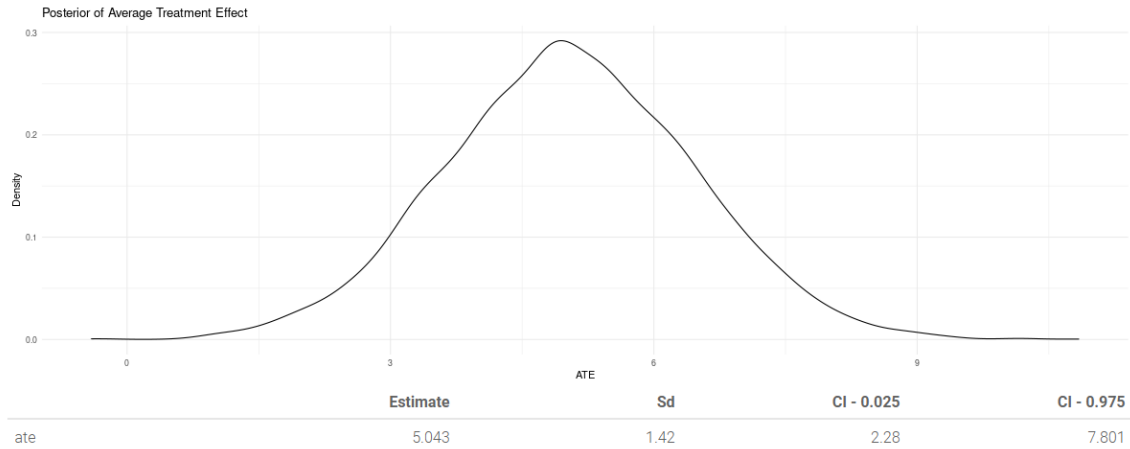


Figure 3. Causal modeling of the number of high-quality comments made



These results provide the answer to our second question (RQ2): providing nudges and visualization cause increased engagement, measured by the number of sessions with AVW-Space. There is also a causal effect of nudges/visualizations on the number of comments written, and more importantly, on the number of high-quality comments. We also found a causal effect on learning, as measured by our conceptual knowledge scores.

3.4 Students' Meetings

In all three years, in the fifth week of the study, students watched and commented on a recording of their own team meeting, while in the following week they rated each others' comments. This allowed us to answer RQ3. Please note that nudges were not given in 2021/2022 while watching the team video. The *Watched* row in Table 7 specifies the number of participants who watched their team video. That number is much lower in 2021, and we believe the reason for that is the COVID-19 lockdown which overlapped with that phase of the study. Table 7 reports the average number of comments on the team video. The majority of students watched their team video in 2020 and in 2022. Although we have previously seen a significant difference in the number of comments on the tutorial/example videos, there is no significant difference on the number of comments on the team video for the three years. This illustrates how interesting students found this part of the study.

Table 7: Engagement with the team videos

	2020	2021	2022
Participants (total)	47	49	47
Watched	43	25	43
Team comments	5.05 (5.65)	4.79 (2.49)	5.54 (3.63)

Students' comments on the team meeting were of different types. Some comments pointed out positive/negative aspects of the team meeting, with students using what they learnt from tutorial videos to critique their own meeting. There were also comments which focused on opportunities for improving their meetings, such as *"Could have better followed an agenda - meeting seems less purposeful."* Students also reflected on the meeting overall (*"I think we had a very effective meeting. We stayed on topic, stuck to the agenda, and left with a good understanding of our next steps."*) or reflected on their own behavior, e.g. *"For myself, if I was to have a*

meeting with someone and they had the same body language as I did in this meeting (at least in the first 10 minutes), I would get an impression that they are not interested, disengaged, or not paying attention.”, “I might be over contributing while not leaving space for others to contribute.” and “I need to construct/deliver my ideas clearer”.

Survey 3 included two questions asking participants about the usefulness of commenting on the team video and rating the comments. We classified the provided feedback manually, based on whether those activities were reported useful or not. The majority of students believed that watching the team video was beneficial (97% in 2020, 70% in 2021 and 88% in 2022). Most responses were about the usefulness of watching themselves, for example *“You can step into an outsiders shoes to view your groups interaction”*. Some feedback expounds it further, pointing out the benefit of watching it from another point-of-view: *“Very useful as an outsider’s perspective allowed me to see things (good and bad) that I did in the meeting more clearly”* and *“You’re able to see your meetings from an outside perspective, which enables you to see things you might’ve missed in the moment such as team members not contributing as much or not paying attention...”*. The responses were reflective, focusing on their own or team member’s shortcomings or mistakes, helping them identify areas for improvement: *“Know my shortcomings and Know how to improve me next time,” “Very useful, it provides me with many of my weakness and my improvements,”* and *“I could communicate to the team what I believe needed to be worked on, and what we were doing well”*.

Most students also found reviewing/rating comments on the team video useful (79% in 2020, 60% in 2021 and 78% in 2022). Similar to the feedback on watching/commenting on the team video, students pointed out that this activity helped them know the perspective of their fellow team members: *“Understanding your team members views, feelings, and opinions, leading to better communication and teamwork”* and *“It not only allowed me to see my thoughts on our team’s communication, but also showed me what the rest of my team thought about how we communicated as well...”*. Some responses pointed out that the anonymized method of reviewing and rating comments was helpful to get or give constructive criticism from/to the team members: *“It was useful to get feedback from my peers in an anonymised forum, which allowed us to make comments that might have been taken defensively if made directly”, “Get some new opinions with no bias on how they think our team communicates and we would receive more constructive criticism”* and *“Forced each team member to consider how others feel and facilitates team growth.”*

To summarize our answer to the third research question (RQ3), we found that the participants were very much interested in watching their own meetings, and reported that they learnt from that activity. One student even pointed out that the feature to watch their own team video was a vital component of AVW-Space, stating that *“It’s a great way to observe how your meetings are structured and how everyone interacts from a third person perspective. I think this aspect of the AVW-Space was its strongest”*.

3.5 Participants’ Perceptions

Our fourth research question focuses on students’ perceptions of the provided training. We address this question in the following subsections, starting with the perceptions of learning as measured by the modified CAP scale. We then analyze the TAM scores, and present the results for the ease of use and perceived usability. The third subsection reports the findings from the NASA-TLX index, while the following subsection presents observations from the last two weeks of the study, when students were watching their own meeting.

3.5.1 Perceptions of Learning

As discussed in Section 2.2, we modified the CAP perception of learning scale [56] to capture students' estimates of learning from their experience with AVW-Space. We first assessed the Cronbach alpha estimate of internal consistency of the three dimensions of the CAP perceived learning scale. The values for the Conceptual ($\alpha = 0.71$) and Affective ($\alpha = 0.61$) dimensions were higher than the value for the Psychomotor dimension ($\alpha = 0.29$). We do not report the alpha value for the whole scale as the scale is not unidimensional [60]. Table 8 reports the averages and standard deviations for the scores on the three dimensions, as well as the overall scores for the 2020 class (no nudges) and the combined 2021 and 2022 classes (when nudges were provided). The scores for dimensions as well as the overall scores are calculated by summing scores on relevant items [56]. The overall score therefore ranges from 0 to a maximum of 48, with higher scores representing higher perceptions of total learning. We performed the independent-samples Kruskal-Wallis test, and found a significant difference on the overall score ($H = 5.23, p = .02$). There was no significant difference on the cognitive scores, but the scores on the affective ($H = 8.49$) and psychomotor ($H = 9.21$) dimensions were significantly different ($p < .005$ in both cases).

Table 8: The scores for the modified CAP perceived learning scale

Nudges	n	Cognitive	Affective	Psychomotor	Overall score
No	30	8.73 (2.18)	11.10 (2.50)	11.27 (1.64)	31.10 (5.18)
Yes	49	8.78 (2.37)	12.69 (3.87)	12.65 (3.25)	34.12 (8.68)

3.5.2 Perceived Usability and Perceived Ease of Use

We integrated ten questions from TAM into Survey 3, in order to assess how much students are likely to use AVW-Space intentionally. Participants provided their responses using the Likert scale ranging from 1 (*extremely likely*) to 7 (*extremely unlikely*). The internal consistency was good, with Cronbach alpha values of 0.95 for perceived usefulness, and .85 for perceived ease of use. Table 9 reports the scores on the two dimensions for the two groups of participants. We found no significant differences on the TAM scores.

Table 9: Scores for the two TAM dimensions

Nudges	n	Perceived usefulness	Perceived ease of use
No	30	3.79 (1.47)	3.53 (1.53)
Yes	49	4.09 (1.40)	3.53 (1.10)

3.5.3 Cognitive Task Index

Survey 3 contained four questions adapted from the NASA-TLX task load index on comment writing, and four questions on comment rating, which allowed participants to specify their perceptions. For each type of activity, participants were asked to specify how mentally demanding it was (*Demand*), how hard they had to work (*Effort*), how discouraged, irritated, stressed and annoyed they felt (*Stress*), and how successful they think they were while performing the activity (*Performance*). Participants provided their responses using the Likert scale ranging from 1 (*Very low*) to 20 (*Very high*). Table 10 presents the scores.

Table 10: Scores for the NASA-TLX index for the two tasks

	2020 (17)	2021/2022 (49)
Comment Writing		
Demand	7.88 (3.90)	8.61 (3.62)
Effort	6.59 (4.02)	8.27 (4.05)
Stress	6.41 (4.87)	7.43 (5.28)
Performance	8.71 (4.57)	11.73 (4.86)
Comment Rating		
Demand	5.05 (3.25)	6.69 (4.30)
Effort	4.29 (2.95)	6.49 (4.62)
Stress	6.05 (4.47)	6.92 (5.84)
Performance	9.65 (5.11)	11.37 (4.52)

The scores from the 2021/2022 participants are generally higher than those of the 2020 students. The only significant difference was found on the Performance when writing comments ($F = 5.05$, $p < .05$, $\eta^2 = .073$), with the 2021/2022 participants reporting more confidence in their performance. For each of the questions, students could provide free-form feedback. Most of the feedback on writing comments stated it was easy to write comments, e.g. *"I just wrote comments when I thought about something during the videos"* but students also pointed out the need to be attentive, e.g. *"It wasn't too mentally demanding, but it did require constantly thinking about when you should add comments about something"*. The participants stated rating comments was neither mentally demanding nor requiring a lot of effort, but also noted that there were too many comments to be rated.

3.5.4 Perceptions of AVW-Space Features

Survey 3 contained open-ended questions which asked participants to specify their opinions of AVW-Space in terms of its functionality. One question was on the usefulness of pausing a video to write a comment. The vast majority of students found this feature useful (96% in 2020, 90% in 2021 and 100% in 2022). Students reported that this feature allowed them to gather their thoughts and focus on writing their comments, for example *"Gives you a chance to think and come up with a proper response rather than moving forward and potentially missing something because you were writing"*. Only two students, one from 2020 and the other from 2021, indicated that pausing the video was not helpful, with only one providing a reason for the answer (*"not very, you lose track"*).

When asked about the usefulness of specifying aspects when composing comments, there was an increase in the number of students who find aspects useful: 46% in 2020, 50% in 2021 and 59% in 2022. Common reasons in the 2022 responses suggest that aspects helped in reflecting on the comments: *"It helps you reflect on the factors you identify in a personal way which helps to connect you with the situation in a neurological sense"*, *"I think that its useful to be able to relate the comment back to an idea"* and *"It gets you to relate what's being discussed to yourself and provokes you to think harder about what aspects of communication are being discussed"*. Interestingly, some students found that aspects are limiting but nevertheless find them useful: *"It's a little bit useful but I feel as though there were not enough possible combinations"* and *"For the most part this was good, however having those as options felt very limiting. Often none of them quite matched and I just had to pick the most relevant one"*. Some students found aspects not useful, as they had other topics in mind or would prefer to use several aspects simultaneously.

In all three years, students appreciated the value of reviewing/rating comments made by peers (83% in 2020, 70% in 2021 and 83% in 2022). The most frequently cited reasons are about being able to see other points of view: *"It is useful as you can see other people's perspective on the videos shown"* and *"You get to see what other people thought of the video and you might get to understand something you didn't before."* Some students also indicated that this activity enabled them to organize their thoughts about the video, and see the ideas or concepts that they

might have missed, for example, “*Extremely useful, allowed me to get other student's perspectives that I might have missed*”, “*You can visualize other people's viewpoints and better allow yourself to learn what you have not found*”. The feedback was generally positive, showing that this feature enabled students to learn from other students and increased their understanding of the concepts discussed in the video. When replying to the question on the usefulness of rating categories when rating comments, many participants noted that they would like to see more rating categories, or even define the categories by themselves.

4 DISCUSSION AND CONCLUSIONS

We presented online training developed for face-to-face communication in software development meetings, and the results from three instances of the second-year software engineering course into which the training was embedded. Our quasi-experimental study allowed us to compare the interactions and learning of the 2020 students, when nudges were not used, to the 2021 and 2022 classes which received the nudges. The nudges AVW-Space provides were designed to encourage constructive behavior, characterized by a high level of engagement in terms of the overall comment writing and also with respect to writing high-quality comments [47].

The findings from the 2020 class are similar to the early studies with AVW-Space [20,45], with 34% of participants passively watching tutorial and example videos. In contrast, there were no passive students in 2021/2022, when nudges were provided. We found significant differences on the various engagement measures between the classes, with 2021/2022 students interacting more with AVW-Space, writing more comments and also writing more high-quality comments. We have also found the causal effects of nudges on the number of sessions with AVW-Space, the number of comments overall and the number of high-quality comments written, as well as the conceptual knowledge score of the target soft skill.

The responses on the modified CAP perceived learning scale show no differences for the Cognitive dimension, but there were significant differences on the scores for the affective and psychomotor dimensions, illustrating participants' attitude towards the skill learnt and also confidence in being able to execute the skill. The analysis of the TAM responses found no significant differences, with all participants finding AVW-Space moderately useful and easy to use. We also found no significant difference on the NASA-TLX responses, with the only exception being that the treatment group reported higher scores for performance on comment writing. Generally students acknowledge the usefulness of commenting on videos, especially their team video. Their responses also confirm that rating of comments is useful, as it enables social learning.

Our study shows that when nudges are given, engagement with tutorial/example videos and learning increase. Interestingly, there were no differences on student engagement with team videos showing how much students were interested in that activity. Watching videos of their own meetings enabled students to reflect on their own contribution to the meeting and how well the team communicated. Many students provided feedback on how useful it was to see their own meeting, which enabled the teams to grow and improve. We found that students felt overwhelmed by the number of comments available for rating, pointing to the need for intelligent filtering for individual students. In future work, we will explore the possibility of providing comments for rating in an adaptive way. One approach for that is to analyze a learner's comments in order to identify weaknesses in terms of concepts mentioned or the depth of reflection, and then to identify relevant comments to present to that student.

Our paper contributes to the literature by showing the effectiveness of active video watching in teaching a new soft skill: face-to-face communication in software development meetings. Furthermore, we show that nudges do

cause increase in interactions with AVW-Space, especially an increase in the number of high-quality comments and learning. We acknowledge that our findings may not be specific to software engineering; however, the chosen soft skill was selected based on its relevance for practicing software engineers. Also, study participants were chosen based on the target audience of our learning platform – software engineering students.

A threat to validity is the effect of the COVID-19 pandemic on the results. These effects were relatively similar in the three years of our study in the country where the research was performed.

There are several limitations of the presented work, the first being the small population size, which is due to the size of the course in which the study was conducted. Additional research is needed to determine the effect of the online training on larger populations of students. We plan to conduct additional studies in the coming years in the same course, which will result in a higher number of participants. The second limitation is that the research was done with students from just one educational institution. Our findings are therefore contextualized and may not necessarily generalize to other tertiary institutions. In future work, we will repeat the study at other institutions, as well as with the industry professionals.

Another limitation is in the way we have assessed students' knowledge. As stated earlier, we used one question in each survey as a proxy for assessing the students' conceptual knowledge of the face-to-face communication skill. This is not ideal for several reasons. Answering surveys can be annoying to students and their responses do not necessarily represent their true knowledge. Furthermore, listing concepts related to the target soft skills does not necessarily mean that the students know how to perform the skill. Ideally, the extent of learning can be assessed by a human expert observing the students in real meetings before and after the training. This requires substantial resources and is not always practically possible. We are currently developing an instrument for self-reporting the ability to communicate face-to-face. However, the instrument is still under development and could not be used in our study. We plan to repeat the study when the instrument has been fully validated.

ACKNOWLEDGMENTS

The work reported in this paper has been supported by the Ministry of Business, Innovation and Employment Smart Ideas grant, as well as by two Ako Aotearoa Southern Hub grants, and a teaching development grant from the University of Canterbury.

We thank Vania Dimitrova for her invaluable contributions to AVW-Space. We also acknowledge the contributions of the members of the Intelligent Computer Tutoring Group at the University of Canterbury, Christchurch, New Zealand.

REFERENCES

- [1] Adam Alami, and Oliver Krancher. 2022. How Scrum adds value to achieving software quality?." *Empirical Software Engineering* 27(7), 165.
- [2] Christina Andersson and Doina Logofatu. 2018. Using cultural heterogeneity to improve soft skills in engineering and computer science education. In *Global Engineering Education Conference*, pp. 191-195.
- [3] Suzanne Anthony and Benjamin Garner. 2016. Teaching soft skills to business students: An analysis of multiple pedagogical methods. *Business and Professional Communication Quarterly*, 79(3), 360–370.
- [4] Lilian Bezard, Marie Debaq, and Astrid Rosso. 2020. The carnivorous yoghurts: A "serious" escape game for stirring labs. *Education for Chemical Engineers* 33, 1-8.
- [5] Charles Bonwell and James Eison. 1991. *Active learning: Creating excitement in the classroom*. 1991 ASHE-ERIC higher education reports. ERIC Clearinghouse on Higher Education, The George Washington University, DC 20036-1183.
- [6] Cynthia J. Brame. 2016. Effective educational videos: Principles and guidelines for maximizing student learning from video content." *CBE—Life Sciences Education* 15(4), es6.

- [7] Alberto Cattaneo, Hans van der Meij, and Florinda Sauli. 2018. An empirical test of three instructional scenarios for hypervideo use in a vocational education lesson. *Computers in the Schools* 35(4), 249-267.
- [8] Mohamed Amine Chatti, Momchil Marinov, Oleksandr Sabov, Ridho Laksono, Zuhra Sofyan, Ahmed Mohamed Fahmy Yousef, and Ulrik Schroeder. 2016. Video annotation and analytics in CourseMapper. *Smart Learning Environments* 3(1), 1-21.
- [9] Michelene T. Chi and Ruth Wylie. 2014. The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist*, 49(4), 219-243.
- [10] Michelene Chi, Joshua Adams, Emily B. Bogusch, Christiana Bruchok, Seokmin Kang, Matthew Lancaster, Roy Levy et al. 2018. Translating the ICAP theory of cognitive engagement into practice." *Cognitive science* 42(6), 1777-1832.
- [11] Hugh A. Chipman, Edward I. George, and Robert E. McCulloch. 2010. BART: Bayesian additive regression trees. *Annals of Applied Statistics*, 4(1), 266-298.
- [12] Orges Cico, Letizia Jaccheri, Anh Nguyen-Duc and He Zhang. 2021. Exploring the intersection between software industry and software engineering education—A systematic mapping of software engineering trends. *Journal of Systems and Software*, vol. 172, 110736.
- [13] Clarke, S.J.; Peel, D.J.; Arnab, S.; Morini, L.; Keegan, H.; Wood, O. EscapED: A Framework for Creating Educational Escape Rooms and Interactive Games to For Higher/Further Education. *Int. J. Serious Games* 2017, 4, 73–86.
- [14] Mauro Coccoli, Ilaria Torre, and Ilenia Galluccio. 2023. User experience evaluation of Edurell interface for video augmentation. *Multimedia Tools and Applications*. 1-23.
- [15] Curtis Conkey, Clint Bowers, Janis Cannon-Bowers and Alicia Sanchez. 2013. Machinima and video-based soft-skills training for frontline healthcare workers. *GAMES FOR HEALTH: Research, Development, and Clinical Applications*, 2(1), 39-43.
- [16] Nuno Correia and Teresa Chambel. 1999. Active video watching using annotation. In *Proceedings of the 7th ACM international conference on Multimedia (Part 2)*, pp. 151-154.
- [17] Michael W. Cronin and Karen A. Cronin. 1992. Recent empirical studies of the pedagogical effects of interactive video instruction in "soft skill" areas. *Journal of Computing in Higher Education*, 3(2), 53.
- [18] Fred D. Davis. 1989. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 319-340.
- [19] Lilian Maria de Souza Almeida. 2019. Understanding Industry's Expectations of Engineering Communication Skills. Doctoral dissertation, Utah State University.
- [20] Vania Dimitrova, Antonija Mitrovic, Alicja Piotrkowicz, Lydia Lau and Amali Weerasinghe. 2017. Using Learning Analytics to Devise Interactive Personalised Nudges for Active Video Watching. In: Bielikova, M., Herder, E., Cena, F., Desmarais, M. (Eds.) *Proc. 25th ACM User Modeling, Adaptation and Personalization conference*, Bratislava, Slovakia, pp. 22-31.
- [21] Vania Dimitrova and Antonija Mitrovic. 2022. Choice Architecture for Nudges to Support Constructive Learning in Active Video Watching. *International Journal on Artificial Intelligence in Education*, 32(4), 892-930.
- [22] Djaouti, D., Alvarez, J., Jessel, J. P., & Rampnoux, O. (2011). Origins of serious games. *Serious games and edutainment applications*, 25-43.
- [23] Samuel Dodson, Ido Roll, Matthew Fong, Dongwook Yoon, Negar M. Harandi and Sidney Fels. 2018. An active viewing framework for video-based learning. In *Proceedings of the Fifth Annual ACM Conference on Learning at Scale*, pp. 1-4. ACM.
- [24] Vincent Dorie, Jennifer Hill, Uri Shalit, Marc Scott, and Dan Cervone. 2019. Automated versus do-it-yourself methods for causal inference: Lessons learned from a data analysis competition. *Statistical Science (with commentary and rejoinder)*, 34(1): 43:68, 94-99.
- [25] Matthias Galster, Antonija Mitrovic and Matthew Gordon. 2018. Toward Enhancing the Training of Software Engineering Students and Professionals Using Active Video Watching. In *Proceedings of 40th International Conference on Software Engineering: Software Engineering Education and Training Track*, Gothenburg, Sweden, pp. 5-8, ACM.
- [26] Matthias Galster, Antonija Mitrovic, Sanna Malinen, and Jay Holland. 2022. What Soft Skills Does the Software Industry* Really* Want? An Exploratory Study of Software Positions in New Zealand. In *Proceedings of the 16th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement*, pp. 272-282.
- [27] Matthias Galster, Antonija Mitrovic, Sanna Malinen, Jay Holland, and Pasan Peiris. 2023. Soft skills required from software professionals in New Zealand. *Information and Software Technology* 160, 107232,
- [28] Silke Geithner and Daniela Menzel. 2016. Effectiveness of learning through experience and reflection in a project management simulation. *Simulation & Gaming* 47(2), 228-256.
- [29] Isabella Graßl, Gordon Fraser, Stefan Trieflinger, and Marco Kuhrmann. 2023. Exposing Software Engineering Students to Stressful Projects: Does Diversity Matter? *Proc. 45th Int. Conf. Software Engineering: Software Engineering Education and Training ICSE-SEET*, 210-222.

- [30] Philip J. Guo, Juho Kim, and Rob Rubin. 2014. 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.
- [31] Sandra G. Hart. 2006. NASA-task load index (NASA-TLX); 20 years later. 50, 904-908.
- [32] Jennifer Hill, Antonio Linero and Jaared Murray. 2020. Bayesian Additive Regression Trees: A Review and Look Forward. *Annual Review of Statistics and Its Application*, 7, 251-278.
- [33] Jennifer Hill, George Perrett and Vincent Dorie. (in press) Machine Learning for Causal Inference. In *Handbook of Multivariate Matching and Weighting for Causal Inference*, ed. Rosenbaum, P., Small, D. and Zubizarreta, J. (accepted for publication in 2022).
- [34] Claudia Iacob and Shamal Faily,. 2019. Exploring the gap between the student expectations and the reality of teamwork in undergraduate software engineering group projects. *Systems and Software*, vol. 157, 110393
- [35] Kenneth R. Koedinger, Jihee Kim, Julianna Zhuxin Jia, Elizabeth A. McLaughlin, and Norman L. Bier. 2015. Learning is not a spectator sport: Doing is better than watching for learning from a MOOC. In Proceedings of the second ACM conference on Learning@ Scale, pp. 111-120.
- [36] Geza Kovacs. 2016. Effects of in-video quizzes on MOOC lecture viewing. In Proceedings of the third ACM conference on Learning@ Scale, pp. 31-40.
- [37] Klaus Krippendorff. 2010. Krippendorff's Alpha. SAGE Publications, Inc. p. 670.
- [38] J. Richard Landis and Gary G. Koch. 1977. The measurement of observer agreement for categorical data. *Biometrics* (1977): 159-174.
- [39] Lydia Lau, Antonija Mitrovic, Amali Weerasinghe, and Vania Dimitrova. 2016. Usability of an Active Video Watching System for Soft Skills Training. Zagreb, Croatia: Proc. 1st Int. workshop on Intelligent Mentoring Systems, Intelligent Tutoring Systems Conference.
- [40] Ze Shi Li, Nowshin Nawar Arony, Kezia Devathasan, and Daniela Damian. 2023. Software is the easy part of Software Engineering - Lessons and Experiences from A Large-Scale, Multi-Team Capstone Course. Proc. 45th Int. Conf. Software Engineering: Software Engineering Education and Training ICSE-SEET, 223-234.
- [41] Ching Liu, Chi-Lan Yang, Joseph Jay Williams, and Hao-Chuan Wang. 2019. Notestruct: Scaffolding note-taking while learning from online videos. In Extended abstracts of the 2019 CHI conference on human factors in computing systems, pp. 1-6.
- [42] Nadia McGowan, Aída López-Serrano, and Daniel Burgos. 2023. Serious Games and Soft Skills in Higher Education: A Case Study of the Design of Compete!. *Electronics* 12(6): 1432.
- [43] Tomas Mikolov, Ilya Sutskever, Kai Chen, Geg S. Corrado & Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. *Advances in neural information processing systems*, 26.
- [44] Negin Mirriahi, Jelena Jovanović, Lisa-Angelique Lim and Jason M. Lodge. 2021. Two sides of the same coin: video annotations and in-video questions for active learning. *Educational Technology Research and Development* 69(5), pp. 2571-2588.
- [45] Antonija Mitrovic, Vania Dimitrova, Lydia Lau, Amali Weerasinghe and Moffat Mathews. 2017. Supporting Constructive Video-Based Learning: Requirements Elicitation from Exploratory Studies. In E. André et al. (Eds.), *Artificial Intelligence in Education*, pp. 224-237.
- [46] Antonija Mitrovic, Matthew Gordon, Alicja Piotrkowicz, and Vania Dimitrova, V. 2019. Investigating the effect of adding nudges to increase engagement in active video watching. In: S. Isotani et al. (Eds.) Proc. 20th Int. Conf. Artificial Intelligence in Education, LNAI 11625, pp. 320-332, Springer Nature Switzerland
- [47] Negar Mohammadhassan, Antonija Mitrovic and Kouros Neshatian. 2022. Investigating the Effect of Nudges for Improving Comment Quality in Active Video Watching. *Computers & Education*, 176, 104340 <https://doi.org/10.1016/j.compedu.2021.104340>
- [48] Negar Mohammadhassan and Antonija Mitrovic. 2022. Investigating the effectiveness of visual learning analytics in Active Video Watching. In: Rodrigo, M.M., Matsuda, N., Cristea, A.I., Dimitrova, V. (Eds.) *Artificial Intelligence in Education, Lecture notes in Computer Science*, 13355, pp. 127-139. Springer, Cham.
- [49] Miguel Hécatl Morales-Trujillo, Matthias Galster, Fabian Gilson, and Moffat Mathews. 2021. A Three-Year Study on Peer Evaluation in a Software Engineering Project Course. *IEEE Transactions on Education* 65(3), 409-418.
- [50] Jeffrey Pennington, Richard Socher & Christopher D. Manning. 2014. Glove: Global vectors for word representation. In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP) (pp. 1532-1543).
- [51] Cyril Picard, Cécile Hardebolle, Roland Tormey & Jürg Schiffmann. 2022. Which professional skills do students learn in engineering team-based projects?. *European Journal of Engineering Education*, 47:2, 314-332.
- [52] Oleksandra Poquet, Lisa Lim, Negin Mirriahi, and Shane Dawson. 2018. Video and learning: a systematic review (2007--2017). In Proceedings of the 8th international conference on learning analytics and knowledge, pp. 151-160.

- [53] Nils Prenner, Jil Klünder and Kurt Schneider. 2018. Making meeting success measurable by participants' feedback. *Proceedings of the 3rd International Workshop on Emotion Awareness in Software Engineering*, 25–31.
- [54] Evan F. Risko, Tom Foulsham, Shane Dawson and Alan Kingstone. 2012. The collaborative lecture annotation system (CLAS): A new TOOL for distributed learning. *IEEE Transactions on Learning Technologies* 6(1), 4-13.
- [55] Vanessa Reher, G. Rehbein, and Peter Reher. 2018. Integrating video recording and self-reflection to enhance communication skills training for dental students. In *International Conference on the Development of Biomedical Engineering in Vietnam* (pp. 715-719). Springer, Singapore.
- [56] Alfred Rovai, Mervyn Wighting, Jason Baker and Linda Grooms. 2009. Development of an instrument to measure perceived cognitive, affective, and psychomotor learning in traditional and virtual classroom higher education settings. *Internet and Higher Education*, 12(1), 7–13.
- [57] Mariecke Schipper, Esther van der Stappen. 2018. Motivation and attitude of computer engineering students toward soft skills. In *Global Engineering Education Conference*, pp. 217-222. IEEE.
- [58] Yvonne Sedelmaier, and Dieter Landes. 2014. Practicing Soft Skills in Software Engineering: A Project-Based Didactical Approach. In *Overcoming Challenges in Software Engineering Education: Delivering Non-Technical Knowledge and Skills*, pp. 161-179. IGI Global, 2014.
- [59] Micael Sousa. 2021. Serious board games: modding existing games for collaborative ideation processes. *International Journal of Serious Games* 8(2), 129-146.
- [60] Keith S. Taber. 2018. The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education*, 48(6), 1273-1296.
- [61] Akio Takashima, Yasuhiro Yamamoto and Kumiyo Nakakoji. 2004. A model and a tool for active watching: Knowledge construction through interacting with video. *Proceedings of INTERACTION: Systems, Practice and Theory*, 331-358.
- [62] Stephen Tang, Martin Hanneghan, and Abdenmour El-Rhalibi. 2009. Introduction to games-based learning. *Games-Based Learning Advancements for Multi-Sensory Human Computer Interfaces: Techniques and effective practices*, 1-17. .New York: IGI Global.
- [63] Ahmed Mohamed Fahmy Yousef, Mohamed Amine Chatti, and Ulrik Schroeder. 2014. The state of video-based learning: A review and future perspectives. *International Journal on Advances in Life Sciences* 6(3), 122-135.
- [64] Meenu Wats and Rakesh Kumar Wats. 2009. Developing soft skills in students. *International Journal of Learning* 15(12).

Appendix

Survey 1

1. What is your age?
 - 18-23
 - 24-29
 - 30-35
 - 36+
2. What is your gender?
 - Male
 - Female
 - Other gender
 - Prefer not to answer
3. What is your first language?
4. How much training have you had on communication in face-to-face meetings?
 - No training
 - Some training
 - Quite a bit
 - A lot
 - Extensive training
5. Select the type of training on communication in face-to-face meetings you have had:
 - Training at high school
 - Training at University
 - Training at community/volunteer group
 - Professional training
 - Other (Please specify)
6. Over the last year, how frequently would you attend face-to-face formal meetings with more than two people?
 - Never
 - Occasionally
 - Once a month
 - Every week
 - Every day
7. Please specify the type(s) of meetings you have had that involved more than two people:
 - Group assignment in high school
 - Group assignment in University
 - Meeting with lecturers
 - As part of an internship
 - Part-time job related to software development
 - Part-time job not related to software development
 - Other (Please specify)

8. How much experience do you have working in software development times outside the university?
- None
 - Some experience (less than a week)
 - Quite a bit (a month)
 - A lot (several months)
 - Extensive experience (more than a year)
9. How often do you watch YouTube?
- Never
 - Occasionally
 - Once a month
 - Every week
 - Every day
10. How often do you use YouTube for learning?
- Never
 - Occasionally
 - Once a month
 - Every week
 - Every day
11. [You have max 3 minutes to answer] Write all words/phrases (one per line) that you associate with effective communication in software engineering meetings.

Survey 2

[You have max 3 minutes to answer] Write all words/phrases (one per line) that you associate with effective communication in software engineering meetings.

Survey 3

[You have max 3 minutes to answer] Write all words/phrases (one per line) that you associate with effective communication in software engineering meetings.

The following questions ask you to estimate how much you have learned from AVW-Space. Please rate, on the scale of 1 (strongly disagree) to 7 (strongly agree), to what extent do you agree with each statement, where lower numbers reflect less agreement and higher numbers reflect more agreement.

- I can summarize what I have learnt in AVW-Space for someone who has not learned from AVW-Space.
- I am able to use the effective meeting participation concepts I learnt in AVW-Space in my future meetings.
- I have changed my attitudes about effective meeting participation as a result of AVW-Space.
- I can assess the quality of face-to-face communication in the example videos used in AVW-Space.
- I feel more confident in my face-to-face communication skills in meetings as a result of AVW-Space.
- I have not expanded my knowledge of effective meeting participation concepts as a result of AVW-Space.
- I can demonstrate to others the effective meeting participation concepts I learnt in AVW-Space.
- I feel that I am a more effective meeting participant as a result of AVW-Space

The following questions have a Likert scale from 1 (Very easy) to 20 (Very hard).

MENTAL DEMAND - Writing comments

How **mentally demanding** was to **write comments on videos** in AVW-Space?

For example, how much mental and perceptual activity was required - thinking, deciding, remembering, looking, searching?

EFFORT - Writing comments

How **hard** did you have to work (mentally and physically) to **write comments on videos** in AVW-Space?

FRUSTRATION - Writing comments

How **discouraged, irritated, stressed and annoyed** did you feel while **writing comments on videos** in AVW-Space?

PERFORMANCE - Writing comments

How successful do you think you were to identify useful points about effective meeting participation when commenting on videos in AVW-Space?

Based on your use of AVW-Space, what would be the usefulness of **pausing a video to write a comment**?

Based on your use of AVW-Space, what would be the usefulness of **asking you to indicate what the comments refer to, e.g. for tutorials: 'I am rather good at this', 'I did/saw this in the past', 'I did not realise I was not doing this', 'I like this point'**;

for examples: 'Verbal communication', 'Feedback', 'Active listening', 'Contribution'.

The following questions have a Likert scale from 1 (Very easy) to 20 (Very hard).

MENTAL DEMAND - Rating comments

How **mentally demanding** was to **review and rate comments on videos** in AVW-Space?

For example, how much mental and perceptual activity was required - thinking, deciding, remembering, looking, searching?

EFFORT - Rating comments

How **hard** did you have to work (mentally and physically) to **review and rate comments on** videos in AVW-Space?

FRUSTRATION - Rating comments

How **discouraged, irritated, stressed and annoyed** did you feel while **reviewing and rating the comments on videos** in AVW-Space?

PERFORMANCE - Rating comments

How **successful** do you think you were to **identify useful points about effective meeting participation** when **reviewing and rating of comments** made by others in AVW-Space?

In the second phase of the study, you experienced two additional features of AVW-Space:

- reviewing the comments on the videos made by other users of AVW-Space;
- rating the comments of other users;

Based on your use of the AVW-Space, what would be the usefulness of **reviewing the comments on the videos made by other people?**

Based on your use of AVW-Space, what would be the usefulness of **rating the comments of your team members on the team meeting video**, e.g. 'This is useful for me'; 'I hadn't thought of this'; 'I did not notice this'; 'I like this point'; 'I do not agree with this'

The AVW-Space system is aimed at **informal learning of soft skills** (e.g. meeting communication, giving presentations, negotiating, managing teams) using selected videos.

The questions below ask how you **perceive the usefulness** of AVW-Space for informal learning of soft skills.

Extremely likely	Quite likely	Slightly likely	Neutral	Slightly unlikely	Quite unlikely	Extremely unlikely
(1)	(2)	(3)	(4)	(5)	(6)	(7)

I think I would like to use AVW-Space frequently.

I would recommend AVW-Space to my friends.

Using AVW-Space would enable me to improve my soft skills quickly.

Using AVW-Space would improve my performance considering the development of soft skills.

Using AVW-Space would enhance my effectiveness when developing soft skills.

I would find AVW-Space useful in my studies/job.

I would find AVW-Space easy to do what I want it to do.

My interaction with AVW-Space would be clear and understandable.

I would find AVW-Space easy to use.

If I am provided the opportunity, I would continue to use AVW-Space for informal learning.

In the last phase of the study, you performed peer-assessment exercise:

Based on the peer-assessment exercise, what would be the usefulness of **watching and commenting on your team video?**

Based on the peer-assessment exercise, what would be the usefulness of **reviewing and rating the comments made by other team members on your team video?**

What do you think is **most exciting** about AVW-Space?

What do you think is **most disappointing** about AVW-Space?