Deep neural networks for collaborative learning analytics: Evaluating team collaborations using student gaze point prediction

Zhang Guo
Computer and Information Sciences, University of Delaware

Roghayeh Barmaki
Computer and Information Sciences, University of Delaware

Automatic assessment and evaluation of team performance during collaborative tasks is key to the research on learning analytics and computer-supported cooperative work. There is growing interest in the use of gaze-oriented cues for evaluating the collaboration and cooperativeness of teams. However, collecting gaze data using eye-trackers is not always feasible due to time and cost constraints. In this paper, we introduce an automated team assessment tool based on gaze points and joint visual attention (JVA) information drawn from computer vision solutions. We evaluated team collaborations in an undergraduate anatomy learning activity ($N = 60$, 30 teams) as a test user study. The results indicate that higher JVA was positively associated with student learning outcomes ($r_{(30)} = 0.50$, $p < 0.005$). Moreover, teams who participated in two experimental groups and used interactive 3D anatomy models, had higher JVA ($F_{(1,28)} = 6.65$, $p < 0.05$) and better knowledge retention ($F_{(1,28)} = 7.56$, $p < 0.05$) than those in the control group. Also, no significant difference was observed based on JVA for different gender compositions of teams. The findings from this work have implications in learning sciences and collaborative computing by providing a novel joint attention-based measure to objectively evaluate team collaboration dynamics.

Implications for practice or policy:
- Student learning outcomes can be improved by receiving constructive feedback about team performances using our gaze-based collaborative learning method.
- Underrepresented and underserved minorities of science, technology, engineering and mathematics disciplines can be engaged in more collaborative problem-solving and team-based learning activities since our method offers a broader reach by automating collaboration assessment process.
- Course leaders can assess the quality of attention and engagement among students and can monitor or assist larger numbers of students simultaneously.

Keywords: collaborative learning analytics, co-located team-based learning, gaze following, joint visual attention, deep learning, experimental design

Introduction

Collaborative learning is an essential educational instrument for teaching and learning. As a team-based and student-centred educational practice, it promotes student motivation and enhances knowledge retention via teamwork and cooperation (Sung & Hwang, 2013). While collaborative learning has been introduced and practiced in co-located settings (Barmaki et al., 2019; Huang et al., 2019; Prinsen et al., 2007; Schneider et al., 2018; Sung & Hwang, 2013), as well as distributed settings (de Freitas & Griffiths, 2007; Li et al., 2008; Schaf et al., 2009; Tao et al., 2019), measuring and evaluating collaboration still remains a challenge. Fairness of group work distribution (Ng et al., 2019), rationality of collaborative conditions (Innes & Booher, 2016) and automatism of process analytics (Rosé et al., 2008) are some of the core issues that need to be considered during collaborative learning analytics, especially in relatively large teams (Bertsimas & Gupta, 2016). Understanding gender effects in collaboration dynamics and investigating best learning practices in teams are also crucial aspects of collaborative learning analytics, which is the focus of our paper.

Gaze-oriented cues can be used as a means of obtaining information about the cognitive activities of a collaborator, and there is evidence that students look at and point to the same object during collaborative

co-located learning activities (Schneider & Pea, 2013; Schneider et al., 2018). This gaze alignment is called joint visual attention (JVA; Van Rheden et al., 2017) – see Figure 1, for example. JVA is a strong predictor of successful collaboration among students (Pietinen et al., 2010; van der Meulen et al., 2016). Compared with collecting traditional self-reported survey data or collaboration system data for one-time performance evaluation (Wendler, 2006), capturing gaze alignment during the entire collaborative process with the JVA features can reveal more valuable information about the quality of interactions among teams (Wahn et al., 2016). New technologies provide innovative methods to extract and measure JVA features. A growing number of researchers have taken advantage of sensor-based eye-trackers to objectively measure gaze features during various social interactions, especially for co-located collaborative tasks (Huang et al., 2019; Schneider et al., 2018). However, despite the provision of highly accurate data from eye-tracking devices, these sensors are usually highly priced and may introduce limitations for educational study settings, specifically those needed to be carried out in classrooms and not in research labs. For example, the calibration might be a time-consuming process or the study activity is needed to be completed within the limited tracking range of the sensors.
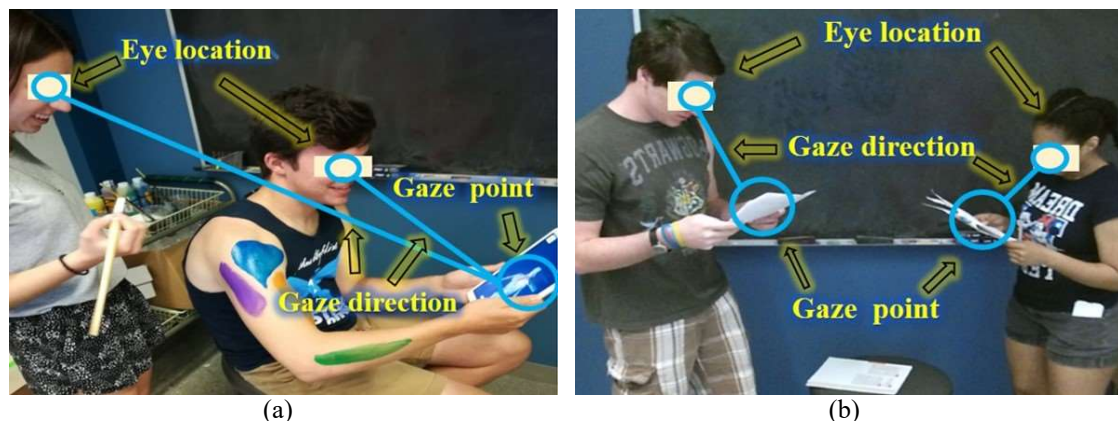

(a)   (b)

Figure 1. Examples of the gaze-following method in our study: (a) with JVA feature: students' gaze points converge on the tablet; (b) without JVA feature: students look at their own notes.

With the emergence of different deep learning techniques, the gaze tracking problem can be approached differently. Using deep neural networks to track the gaze features from a sequence of 2D images or videos is practical and robust in understanding and interpreting student behaviours in human-human and human-object interaction (Lian et al., 2018; Recasens et al., 2015). For example, when two students are looking for a path from the library to the gym on a single university map, by following their gaze direction, we can easily find out if they are sharing the same information and we can predict whether they will pick the same path. Compared with eye tracking and gaze estimation, the gaze following method (Hansen & Ji, 2009; Lian et al., 2018; Santini et al., 2017) not only estimates the gaze direction but also predicts the gaze point from the image without the need for specialised hardware (e.g., head-mounted camera, infrared light source) and obtrusive gaze calibration procedure. Figure 1 shows the application of gaze following method on top of two images captured from our test user study.

In this paper, we introduce a computer vision-based solution for team performance evaluation using mutual gaze point predictions, along with a collaborative anatomy learning activity as a test user study for our approach. We recorded collaborative activity sessions as a sequence of images with a colour camera. For collaboration analysis, we first tracked team members' gaze directions and the focus objects during the activity using the gaze following method (Lian et al., 2018) with a deep neural network framework. We then extracted the JVA features of the teams and analysed them with other collected data, including post-test scores and demographics information related to team gender composition. This study hypothesised that students who share mutual gazes during the activity – for example, those teams with higher JVA values – obtain higher scores in their post-activity knowledge tests as well since they engage more in collaborative tasks. We were also interested in understanding if JVA values are varied significantly in different study conditions and different gender compositions of teams for collaborative learning.

This paper is organised as follows: First, we review the literature on the related work in anatomy education, gender effects, collaborative learning analytics, JVA applications and gaze following approaches. We then introduce details of the intervention, data sets, the gaze following method and our proposed assessment measures for collaboration. Finally, we present the findings from our test user study and discuss further implications.

Literature review

Educational technology for anatomy education

In the domain of human anatomy learning, different education technologies have recently replaced traditional teaching methods such as lectures, cadavers and textbooks. With modern computer-assisted technologies, 3D visualisation methods improve students' performance by allowing them to explore 3D anatomical models on 2D mobile device screens (Yammine & Violato, 2015). Mobile-based applications and web-based 3D games have been used as learning tools for the study of human skeletal, muscular and cardiovascular systems to name a few (Golenhofen et al., 2020; Lemos et al., 2019). Virtual reality (VR) and augmented reality (AR) techniques have been adopted into medical education and surgical training fields in recent years (Maresky et al., 2019; Silva et al., 2018). As powerful learning tools, VR and AR engage students in an immersive environment with audio and visual interactions, and stereoscopic 3D models to enhance their learning experience (Barmaki et al., 2019; Hackett & Proctor, 2018; Luursema et al., 2006, 2008). We evaluated team performance in a controlled study that leveraged modern anatomical content visualisation in 3D with handheld tablet devices and large-scale AR displays.

Gender effects and collaboration

As the results of the different socialisation processes, gender differences have been discussed by recent studies at the team level. The relationship between gender and collaboration is not uniform, and it varies based on different disciplines and tasks (Fernandez-Sanz & Misra, 2012; Wegge et al., 2008). Researchers have shown that females have better information-processing skills than males during cognitive tests (Rabbitt et al., 1995; Schaie & Willis, 1993). Females' higher management ability in collaborative tasks has also been highlighted (Bear & Woolley, 2011; De Paola et al., 2018; Eagly & Carli, 2003). Several studies (Bear & Woolley, 2011; De Paola et al., 2018; Eagly & Carli, 2003; Rabbitt et al., 1995; Schaie & Willis, 1993) concluded that collaboration performance would be improved with females involved. Other research showed that women often had negative experiences on teams due to gender biases at the technical level, especially in science, technology, engineering and mathematics (Meadows et al., 2015; Meiksins et al., 2015). Conversely, some reported no significant gender effect. Andersson (2001) argued that, although females have better performance on individual memory tasks, no gender effect was found in collaborative tasks. Prinsen et al. (2007) noted that females were more likely to collaborate, and males were more assertive in computer-mediated communication (Herring, 1996), and computer-supported collaborative learning (CSCL; Lipponen, 1999) settings. In the same 2007 study, Prinsen et al. also acknowledged that different distributions of roles in collaborative learning might change gender contributions. We explored potential gender differences in our anatomy learning study in association with learning outcomes and joint attention measures.

Collaborative learning analytics

The importance of social interactions during the learning process has been emphasised (Okita et al., 2008). Collaborative learning not only helps students to improve teamwork skills but also promotes learning motivation, increases learning experience, enhances brainstorming skills (Guo & Barmaki, 2019; Webb et al., 1995) and facilitates their learning performance during team interactions (Barmaki et al., 2019; Sung & Hwang, 2013). Consequently, researchers have highlighted that instead of using new learning formats, more attention should be paid to the measurements of and access to collaboration performance (Huang et al., 2019). In early attempts to analyse collaborative learning, Soller and Lesgold (1999) provided a practical collaborative learning framework and evaluated active learning skills using conversational interaction data collected from surveys. With advancements in CSCL research, machine learning techniques have been used to predict student grades, using support vector machines (Baker & Yacef, 2009), decision trees (Rosé et al., 2008) and regression (Kotsiantis, 2012) to name a few. Those solutions either established effective collaborative learning models or built reasonable standards for evaluating collaboration performance based

on single-time solicitation techniques. However, the data used in those models were collected from class attendance, quiz scores, or reports, which may represent only students' one-time or episodic performance during the learning activities.

Collaborative learning analytics research has been much of attention mainly in flipped classrooms with collaborative problem-solving activities primarily in science, technology, engineering and mathematics, mediated by computers either in co-located or distributed learning settings. The majority of research for distributed studies lies in CSCL studies; for example, Subburaj et al. (2020) presented a collaborative problem-solving model for an educational physics game with 101 teams of undergraduate students. Facial expressions, acoustic-prosodics, eye gaze and task context information were captured in the last minute of the intervention and used as measures for predicting success at solving the game. The combined predictive model of non-verbal cues with language-based features outperformed other predictive models. Also, behavioural cues such as eye gaze (Vrzakova et al., 2019), head pose (Otsuka et al., 2018), prosody and acoustics (Miura & Okada, 2019; Murray & Oertel, 2018), as well as language (Flor et al., 2016) have been investigated in collaborative learning analytics for group outcomes including task performance. In co-located learning scenarios, despite their similarity of approaches to distributed settings, physical proximity and movement dynamics in teams were a key factor in the collaboration. In our previous work (Guo & Barmaki, 2019), we used an object detection approach atop image data from a collaborative anatomy learning activity to extract useful proximity features, such as the locations of students and objects in the scene. Research also used multimodal learning analytics techniques and high-level features from dissimilar sources such as video and sketchpads to discriminate between experts and non-experts in groups of students (Ochoa et al., 2013), and understand team performance from physical engagement, satisfaction and individual accountability perspectives (Spikol et al., 2017).

JVA applications

JVA features have been introduced to a broad range of applications, including collaborative search (Brennan et al., 2008), mediated interaction (Bente et al., 2007), infant-caregiver interaction (Markus et al., 2000) and training for children with autism (Whalen & Schreibman, 2003). Interest has grown in the use of synchronised eye-trackers to quantitatively measure gaze alignment in various collaborative situations (Bryant et al., 2019; Huang et al., 2019; Kim et al., 2020; Van Rheden et al., 2017). However, there are challenges in using eye-tracking sensors, including the high cost of the devices, and restricted environmental and calibration settings (e.g., the camera should be precisely in front of the student within a close distance and on top of a specific panel (Huang et al., 2019)). Image-based computer vision methods – as a more affordable alternative approach – have also been used for extracting gaze features in previous studies. Using a colour camera, Yücel et al. (2013) presented an image-based head pose estimation method for establishing joint attention between an experimenter and a robot. Harari et al. (2018) used image segmentation to identify the common gaze target by combining the estimated 3D gaze direction.

Gaze following using deep learning

There has been an expanding interest in the estimation and reconstruction of human gaze direction from 2D images to identify their activities in the scene using various deep learning frameworks. Gaze following is the task of following people's gaze in a scene and inferring what they are looking at. Compared with eye-tracking and gaze estimation, gaze following not only estimates the gaze direction but also detects the focus point from the image (Lian et al., 2018). Patacchiola and Cangelosi (2017) proposed a face detector to extract face landmarks and estimate head poses using convolutional neural networks. Marín-Jiménez et al. (2014) used head pose detection with implicit pose information to detect human-human interaction in videos. However, those works were limited by the complexity of inputs (massive eye-tracking data: (Yücel et al., 2013); restricted situations (resolution of the image: (Marín-Jiménez et al., 2014); and field of view (the distance between the camera and students: (Zhu & Ramanan, 2012). In the work of Recasens et al. (2015), the gaze point of multiple observers in daily scenarios was predicted using deep neural networks and saliency models of attention. Mukherjee and Robertson (2015) combined RGB-depth images and multimodal data to reconstruct 3D head poses and follow gaze direction in images and videos. These studies motivated the work reported here, to use a deep learning approach to target gaze alignment features for the novel application of collaborative learning analytics.

We were interested in understanding how two students are interacting with one another, or with objects, and following the gazes of multiple observers in a scene. During the preparation stage of our study, we tested various algorithms for gaze feature extraction, including facial landmark detection (Patacchiola & Cangelosi, 2017) and head pose detection (Marín-Jiménez et al., 2014) to predict gaze direction. However, neither facial landmark nor head pose can be detected completely when participants are back to the camera or face downwards, which was not practical for our study (see Figure 2(a) and (b)). Hence, we used gaze following method (Lian et al., 2018) to estimate both the gaze direction and the gaze points to collect human-human interaction information. Further details about our approach are presented in the following section.
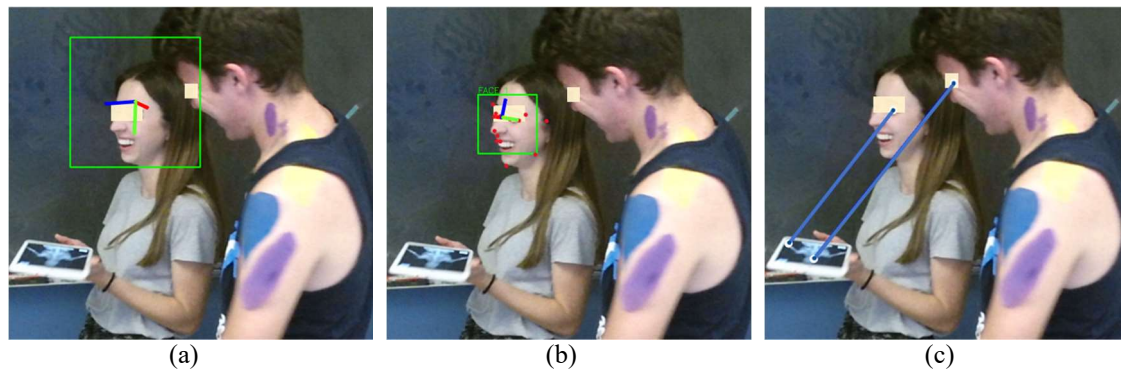

| (a) | (b) | (c) |

Figure 2. Example results from the application of multiple methods on our data set (a) head pose detection (Marín-Jiménez et al., 2014), (b) facial landmark detection (Patacchiola & Cangelosi, 2017) and (c) gaze following method (Lian et al., 2018).

Method and materials

Intervention

We conducted a between-subjects study of collaborative muscle learning intervention in a laboratory course (General Biology) as part of an undergraduate premedical program at Johns Hopkins University. A total of 301 students in 138 teams participated in the original study, we selected the data from a subset of teams with two members ($N = 60$, 30 teams) as our test data set. Students worked in teams to complete a muscle painting activity (Barmaki et al., 2019; Marieb, 2015) as part of their required laboratory activities. They were expected to identify and paint the major muscles of their body using one of the learning instruments (textbook, tablet or AR) and washable painting supplies. The first student played the role of a model, while their teammate, as a painter, located the major upper-limb muscles with the aid of their laboratory manual (Marieb, 2015) or other digital devices and painted the model's upper limb. Afterward, students switched roles, and the upper-limb painter became a model for the lower limb. The goal of the learning activity was to ensure all students could gain knowledge of anatomy in a collaborative effort. See Figure 3 to learn more about the intervention details.

As briefly mentioned, our study had three different settings based on instrumental tools. Students in the control group used textbooks as their learning tools. In experimental group I, instead of a textbook, students used our in-house interactive app on the tablet as a 3D musculoskeletal visualising system. Experimental group II used a screen-based AR system – also developed internally – where students could see themselves with augmented anatomy visualisations on a large display (Barmaki et al., 2019). The knowledge base information presented in all instrumental tools was identical to mitigate potential confounding factors related to student workload and learning. There was also a mobile workstation inside the laboratory room to capture snapshots from students during learning activities. Figure 3 shows the three study conditions of the learning activity.
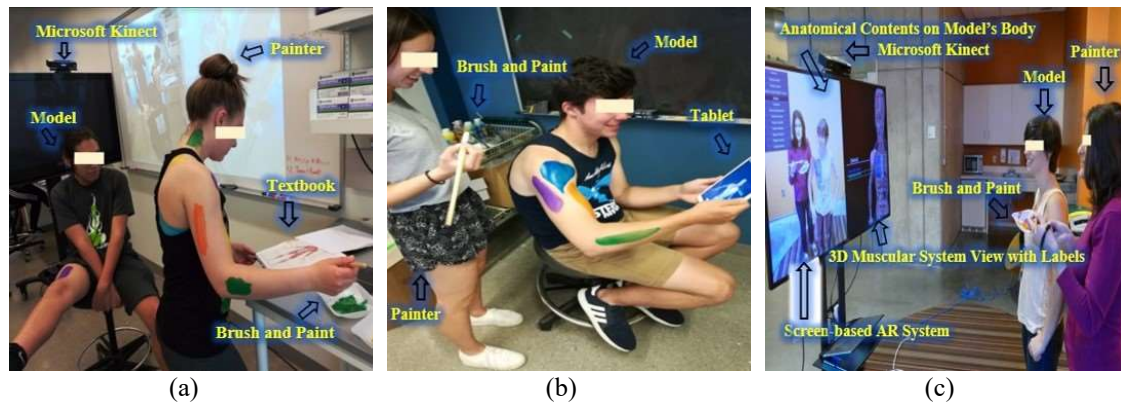
(a)   (b)   (c)

Figure 3. Study conditions for students in pairs to complete anatomy painting intervention using (a) a textbook, (b) an interactive app on the tablet, or (c) a screen-based AR system.

The study was approved by the Institutional Review Board (Protocol ## HIRB00005021) in May 2018, and oral informed consent was obtained from each participant student before the study commenced. After consent, students entered the activity room with their teammates and completed the task. All students completed both pre- and post-activity questionnaires and knowledge tests.

Data sets

Surveys
Using the Qualtrics application, survey data was collected from all students individually after completing the activity. The survey consisted of demographics information, usability questions and a post-test about the human muscle system.

Image training data
We adopted GazeFollow, the large-scale gaze-following data set used for training from Recasens et al.'s (2015) study. This benchmark data set had 130,339 people and 122,143 images in total with gaze points inside the image.

Image test data
The test data set consisted of 4,646 images collected from 30 teams (pairs) of students during the collaborative learning activity in the three conditions (10 teams from each condition of the textbook, tablet and AR, totalling images from 30 teams). Images were captured every 10 seconds, and each image file was timestamped. The resolution of each test image was 2560 x 1440 pixels. Images with camera difficulties or additional individuals in the scene were discarded.

Gaze following framework

To extract shared gaze features from the images, we needed to estimate the students' gaze direction and focus in the scene. Thus, we applied a two-stage gaze following approach (Lian et al., 2018) on our test data set. This method was very suitable for our project since it was capable of detecting the gaze direction from the head image and predict the potential gaze point along the gaze direction, via deep neural networks. The gaze following approach and its underlying network architecture is shown in Figure 4.

The gaze following framework was inspired by the human behaviour of gaze following (Lian et al., 2018). First, a gaze direction was estimated from the gaze direction pathway. In the gaze direction pathway, the resized head image (224 x 224) – image sizes are listed in pixels hereafter – was fed into the convolutional neural network ResNet-50 (Rezende et al., 2017) for feature extraction. Then, the head features were concatenated with head position features encoded by one fully connected layer for gaze direction estimation. A coarse gaze direction was predicted as the vector output and then encoded as multi-scale gaze direction fields. The gaze point was assumed to be in the gaze direction or line of sight. Next, the multi-scale gaze direction fields were combined with the scene contents (224 x 224) and fed into the heat map pathway for heat map regression using a feature pyramid network (Zhao et al., 2019).

The heat map (56 x 56) represented the probability distribution of the gaze point, and the point with the maximum value of the heat map represented the probable gaze point of the scene.
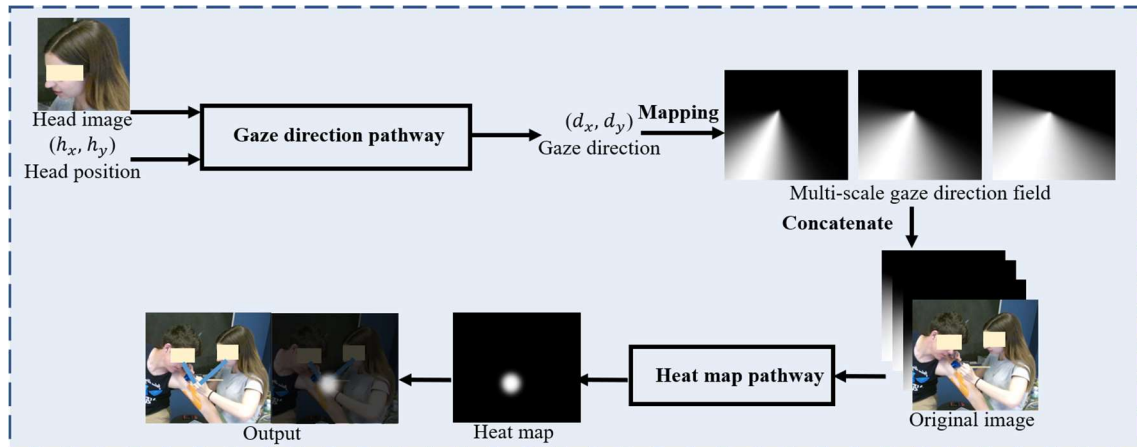

Figure 4. The network architecture for the gaze following method (Lian et al., 2018) atop our collaborative study image frames. Using the heat map, we can predict the gaze point convergence (focus point) of students in the collaborative activity.
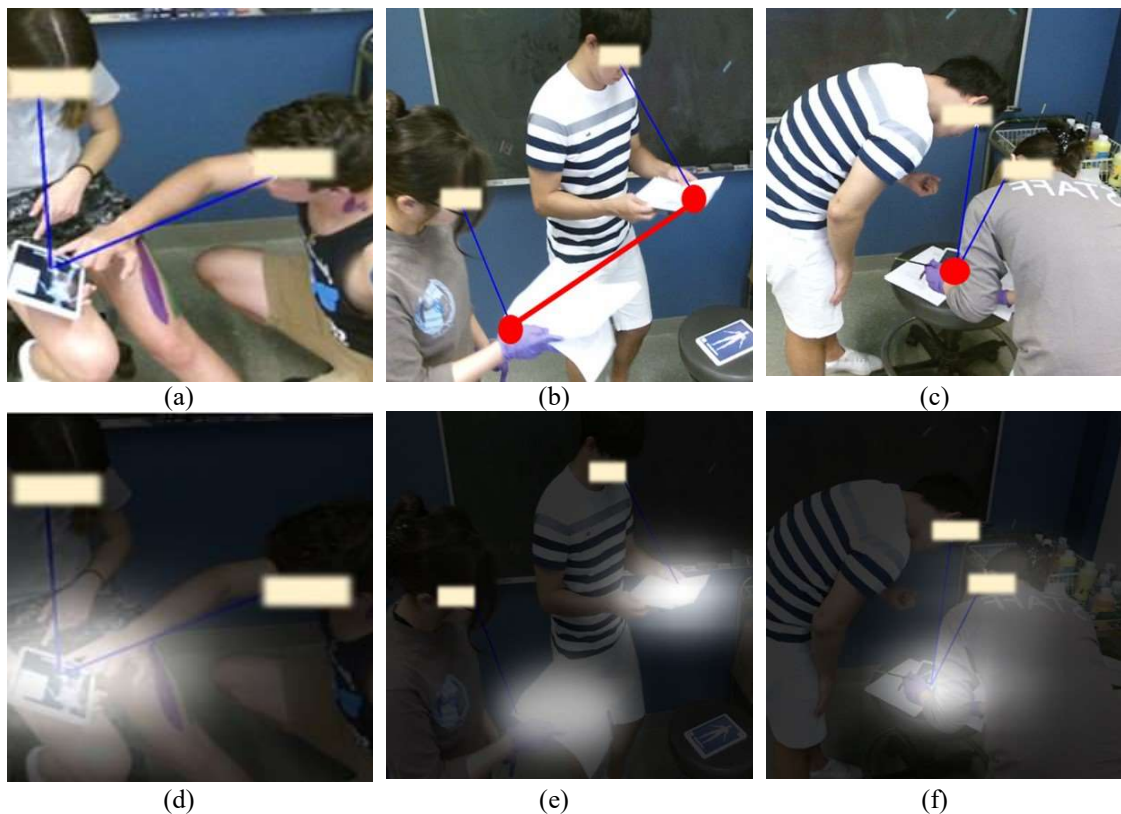

Figure 5. Gaze following results for three sample frames: (a) gaze directions with blue lines; (b) output without the JVA feature (Euclidean distance between the gaze points of students is greater than 100 pixels; (c) output with the JVA feature (Euclidean distance between the gaze points is smaller than 100 pixels); and (d–f) heat maps associated with the gaze points.
Note. While two distinct sheets of papers are predicted as gaze points for team members in (e), (d) and (f) are examples of the convergence of visual attention on the tablet device; thus, the Is_JVA variable is true in (d) and (f).

Lian et al. (2018) claimed that their gaze following approach outperformed other existing methods in gaze point prediction. Compared with state of the art (Recasens et al., 2015), Lian et al.'s method decreased 23.68% of the Euclidean distance error for gaze point on the GazeFollow data set. We chose Lian et al.'s

gaze following model because it managed to simulate the gaze following behaviour of a third person view. Furthermore, Lian et al. trained this model robustly on a large data set by using the heat maps for focus point prediction. Figure 2 highlights its strengths over other existing solutions.

The gaze following output shown in Figure 5(a) visually draws a blue gaze line on the original image for each individual in the scene. The blue line is initiated at the eye location and terminated at the predicted final gaze point. The highlighted region in the corresponding heat map – Figure 5(d–f) – represents the predicted gaze point where the students are looking. The output also marks the coordinates of each gaze point – which are used in our approach as a collaboration metric. We were interested in the automatic recognition of joint or mutual gaze visual attention among students in every image sequence during the collaborative task. Further details about the JVA feature analysis as a collaboration measure are presented in the following section.

Evaluation measures

We analysed team performance and collaboration based on objective measures related to joint attention, knowledge retention, study conditions and gender composition of the teams. These evaluation measures of collaboration are described as follows.

JVA ratio
JVA represents the shared focus of two or more individuals and plays a key role in collaboration prediction (Bruinsma et al., 2004). In this work, we defined the JVA ratio for each team based on the frequency with which the two students shared gazes during the collaborative activity, divided by total image frames captured from the team – a normalised measure of JVA based on the total activity frames of the teams. Since there was a lot of cooperation between painters and models during the learning activity process, they needed to maintain joint attention most of the time: while painting, discussing and looking at the learning materials. For example, when the painter was painting, both painter and model may have looked at the same location, the active painting region. When students needed to find the muscle's correct location, the painter and the model may have shared the screen of the interactive app on the tablet or the AR view to zoom in on the 3D musculoskeletal system.

We used Euclidean distance between the gaze points detected by the gaze following method for automatic identification of the JVA in each image frame (Is_JVA was a Boolean variable per each frame; it was set to false by default). Based on image size and resolution, we recognised the JVA - our mutual gaze feature - and set Is_JVA as true, if that distance was smaller than 100 pixels. For each team, the JVA ratio was computed by number of frames in which the Is_JVA variable was true, divided by the total number of frames. Some examples with and without JVA recognition are shown in Figures 1 and 5. We were interested to learn if a higher JVA ratio is associated with better learning outcomes.

Team post-test score
The key objective of this collaborative learning activity was to enhance the anatomy knowledge retention of students. All participants needed to independently – not with the assistance of their peers – locate and label five muscle names in a diagram of the human musculature in the post-test; thus, individual test scores ranged between 0 and 5 with discrete values. Since we used the average of post-test scores per team and named it team post-test score, the team post-test score was still in the same range, but non-discrete values were also observed in the data set.

Study conditions
As mentioned earlier, there were three different conditions or settings for our muscle painting study. Students in the control group used textbooks as their learning tools. The experimental groups either used a tablet or the screen-based AR system to complete the task. We wanted to investigate the differences in team performance based on these three conditions and the two groups.

Gender composition
Students were preassigned randomly to teams to complete the muscle painting activity. There were three possible gender compositions per pair of students: male pair, female pair and mixed pair. We were interested in evaluating the gender effects in collaborative learning and investigating if any significant variability of JVA ratios and knowledge tests was present in female-female, male-male and male-female

(mixed) pairs of students. As females had a higher enrolment rate in the General Biology lab – which was a common pattern in premedical programs across the nation (Barzansky, 1997; Brooks, 2017) – our study participants were also mostly females with a total of 38 out of 60 study participants (out of the 30 teams, half of them were female-only, eight were mixed and the remaining seven were male-only teams). This does not mean gender imbalance in the data since the sample population is a representative subset of the target population in premedical programs (Barzansky, 1997; Brooks, 2017).

Results

In the following, we report descriptive and inferential results from our test user study. In particular, we looked at the JVA ratio – an automatically generated measure based on our proposed framework using deep neural networks – in association with our evaluation measures. Table 1 summarises the descriptive statistics for pairs of students in each study condition, including the number of teams, mean values and standard deviations for JVA ratios and team post-test scores.

Table 1
Summary of JVA ratio and team post-test score with different instrumental tools

Group	Condition (Instrumental tool)	Observation (teams)	JVA ratio (%)	Team post-test score
		n	$M \pm SD$	$M \pm SD$
Control	Textbook	10	31.30 ± 9.73	1.15 ± 0.95
	Tablet	10	46.50 ± 15.43	2.35 ± 1.03
	AR	10	44.60 ± 17.28	2.35 ± 1.42
Experiment	Combined (tablet & AR)	20	45.55 ± 15.97	2.35 ± 1.20
Total	Textbook, tablet, AR	30	40.80 ± 15.59	1.95 ± 1.25

Note. JVA = joint visual attention; M = mean; n = number of observations; SD = standard deviation; Team post-test score is in range from 0 to 5; JVA ratio (%) is in [0–100] range.

Participants

We analysed data from 60 participants (38 females) in 30 teams. All of these students were enrolled in the undergraduate premedical program at Johns Hopkins University. There were 10 teams for each condition – textbook, tablet and AR. Knowing that tablet and AR conditions were part of the experimental group, we had 20 teams in the experimental group and 10 in the control group. Data from teams with a larger size, those in different activity rooms, those with students under 18 years of age and those with incomplete data were excluded in this study.

JVA ratio

JVA ratio was the percentage of the time teams had shared mutual or joint attention during the learning activity. Although no significant difference between the three study conditions and the JVA ratio was observed, the p value was very close to the critical value of α ($F_{(2,27)} = 3.26$, $p = 0.054$, *ns* – *ns* stands for statistically non-significant). Interestingly, the JVA ratio of the two experimental groups of tablet and AR ($n = 20$, $M = 45.6$, $SD = 15.97$) was higher than those in the control condition who used textbook ($n = 10$, $M = 31.3$, $SD = 9.73$) and this finding was statistically significant with a large effect size ($F_{(1,28)} = 6.65$, $p < 0.05$, Cohen's $d = 1.00$). Table 1 and Figure 6 provide additional information about the JVA ratio distribution across all study conditions and groups.
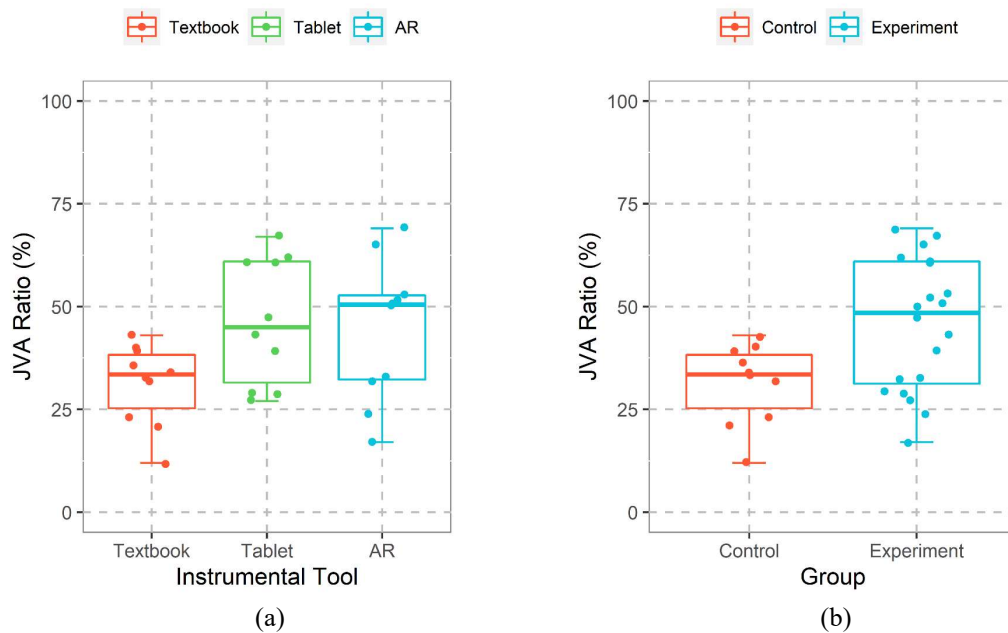
(a)   (b)

Figure 6. The boxplot with observed data points for JVA ratio across (a) different study conditions (instrumental tools) of textbook, tablet and AR, (b) two groups of control (textbook) and experiment (tablet and AR). JVA ratio was significantly different between control and experimental groups.

Team post-test score

A significant difference based on team post-test scores was observed among study conditions of textbook ($M = 1.15$, $SD = 0.95$), tablet ($M = 2.35$, $SD = 1.03$) and AR ($M = 2.35$, $SD = 1.20$), ($F_{(2,27)} = 3.64$, $p < 0.05$, $r^2 = 0.16$, medium effect size). Post-hoc comparisons indicate that pairs of textbook and tablet, and textbook and AR conditions were different from each other based on differences in the means. Similarly, the team's average post-test score from the two experimental groups of tablet and AR ($M = 2.35$, $SD = 1.20$) was significantly higher than those in the control group ($M = 1.15$, $SD = 0.95$), and this finding was also statistically significant with a large effect size ($F_{(1,28)} = 7.56$, $p < 0.05$, Cohen's $d = 1.06$). See Figure 7 to learn more.

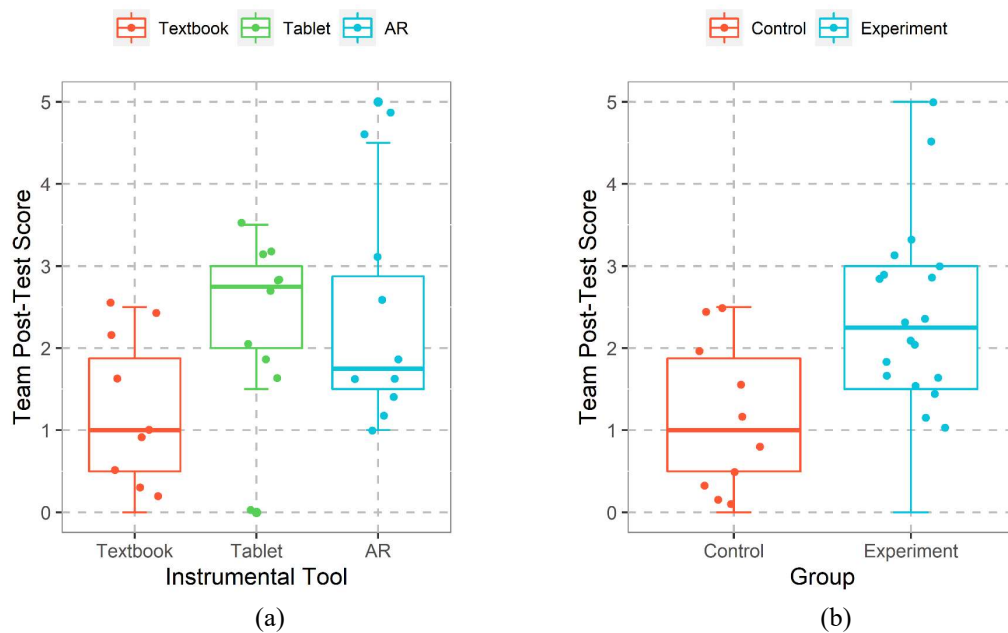(a)   (b)

Figure 7. The boxplot with observed data points for team post-test score across (a) different study conditions (instrumental tools) of textbook, tablet and AR, (b) different groups of control and experiment.

JVA ratio and team post-test score

We also measured the association between JVA and team post-test scores using the Pearson correlation coefficient. The Pearson correlation measure indicates a significant positive linear association (Zou et al., 2003) with a strong relationship between the JVA ratio and team post-test scores ($r_{(30)} = 0.50$, $F_{(1,28)} = 9.33$, $p < 0.005$, $r^2 = 0.25$, large effect size). The scatter plot drawn from data is shown in Figure 8. This finding shows that JVA features are strongly associated with learning outcomes, such as post scores. Points on the scatter plot closely resemble a straight line with a positive slope, which shows that post-test scores increase with higher JVA ratios. Therefore, the team with a high frequency of sharing gazes is more likely to achieve better outcomes in the post-test.
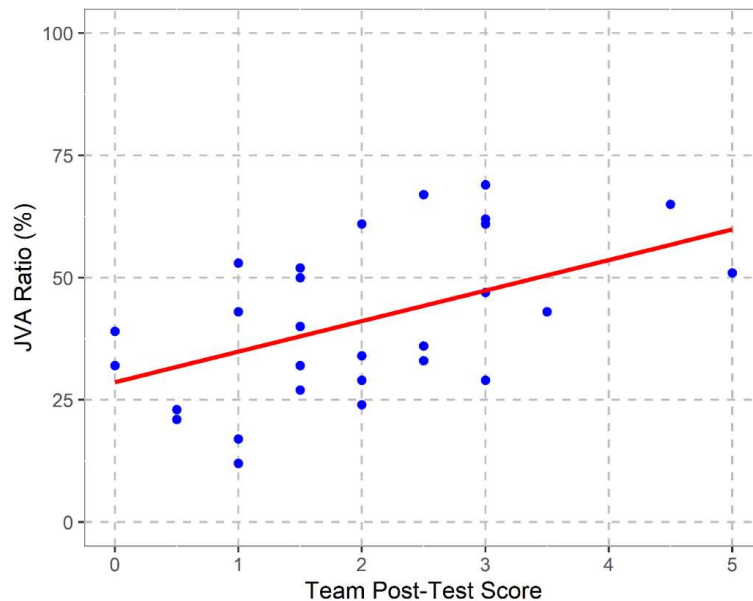

Figure 8. The scatter plot of JVA ratio with team post-test scores. The Pearson correlation and its underlying regression model indicate a significant positive correlation between JVA ratio and team post-test score.

Gender composition

We recorded the gender composition for each of the 30 teams from the survey data and investigated the gender effects on collaborative learning during the activity (see Table 2 and Figure 9 to learn more). Overall, mixed pairs (eight teams) achieved the highest JVA ratio ($M = 47.0$, $SD = 15.46$) and the best learning outcomes from post-test scores ($M = 2.50$, $SD = 1.28$), but this variability was not statistically significant ($F_{(2,27)} = 1.29$, $p = 0.29$, *ns*). Moreover, no significant difference was observed based on gender composition in teams for JVA ratios ($F_{(2,27)} = 1.10$, $p = 0.35$, *ns*).

Table 2
Summary of JVA ratio and team post-test score with different gender compositions

Gender composition	Observation (teams)	JVA ratio (%)	Team post-test score
	n	*M ± SD*	*M ± SD*
Females	15	37.00 ± 15.05	1.63 ± 1.29
Males	7	41.86 ± 16.72	2.00 ± 1.04
Mixed	8	47.00 ± 15.46	2.50 ± 1.28
Total	30	40.80 ± 15.59	1.95 ± 1.25

Note. JVA = joint visual attention; *M* = mean; *n* = number of observations; *SD* = standard deviation; Team post-test score is in range from 0 to 5; JVA ratio (%) is in [0–100] range.
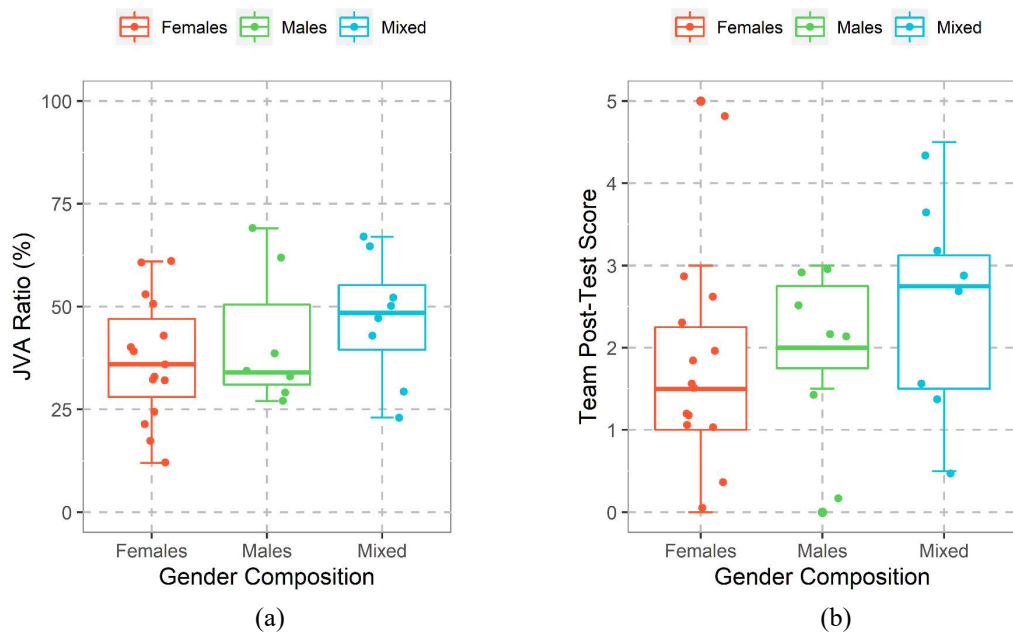
Figure 9. The boxplot with observed data points across teams with different gender compositions: (a) JVA ratios, (b) team post-test scores. No significant difference was observed in the study for JVA ratios nor post-test scores for different pairs of students.
Note. Among these 30 pairs or teams of participants, there were 15 female-female, seven male-male and eight mixed pairs.

Discussion

JVA ratio

Capturing gaze alignment during the collaborative process with the JVA criterion can reveal valuable information about the quality of interaction among teams (Bruinsma et al., 2004; Bryant et al., 2019; Markus et al., 2000; van der Meulen et al., 2016; Wahn et al., 2016); although, not many studies have investigated computer vision–based approaches to better measure and capture it in co-located team-based learning activities. In this paper, we introduced a novel assessment tool for automatic team performance evaluation using mutual gaze information using the gaze following method (Lian et al., 2018). Compared with other methods using traditional one-time performance evaluation (Wendler, 2006) or high-cost eye-tracking devices (Bryant et al., 2019), our method was able to automatically extract JVA features during the whole learning process with a simple colour camera. We also investigated the effectiveness of our JVA method in a test user study. Results show that the JVA ratios of the two experimental groups of tablet and AR were significantly higher than those in the control group, who used the textbook. Our findings are supported by research based on gaze information from student users that looked at e-textbooks as a potential alternative learning tool (Gelderblom et al., 2019), although that research was limited to individual learners and not teams.

Team post-test score

Post-test scores indicate student achievement from the learning activity (Morris, 2008). In this study, we set up three different study conditions by using different instrumental tools for an anatomy learning activity. Right after the activity, post-test scores were collected from students using a survey completed individually, and the team post-test score was calculated as the average of team members' individual test scores. Team post-test scores of the two experimental groups of tablet and AR were significantly higher than those in the control group. Furthermore, our research on collaborative learning analytics conducted with 288 students in May 2017 (Barmaki et al., 2019) also showed that higher test scores were achieved from experimental groups who used the AR system. These findings are in agreement with previous studies in anatomy education, which highlighted the potential of using evolving technologies such as mixed and augmented

realities for enhancing student learning and outcomes in anatomical science education (Maresky et al., 2019; Nicholson et al., 2006; Silva et al., 2018).

JVA ratio and team post-test score

There was a significant positive linear association with a strong relationship between the JVA ratio and the team post-test score. This is in agreement with the hypothesis of our study: students who shared mutual gaze with their teammates for a longer time on the learning task were more likely to obtain higher scores in their post-activity knowledge tests. This finding also agrees with previous research on the positive effects of sharing a gaze on learning outcomes, including imitation and socio-cognitive performance (Carpenter & Tomasello, 1995; Hirotani et al., 2009). Features like hand distance, speed and face count have been used as high-level features for collaborative learning analytics (Spikol et al., 2017), and they seem to have potential in practice-based learning. Our findings from the JVA and team performance outcomes also show the potential of proximity-based, behavioural measurements in co-located or practice-based collaborative learning analytics.

Gender composition

There were three possible gender compositions per pair of students: females, males and mixed pairs. About half of the participants were in female-only teams, which is also a common gender enrolment rate in life sciences and premedical programs (Barzansky, 1997; Brooks, 2017). Even though mixed teams had slightly higher JVA ratios and better learning outcomes, no significant difference was observed based on the gender composition of the teams neither for the JVA ratio nor for the team post-test score. These findings are consistent with previous studies that noted no gender effect in life sciences studies (Andersson, 2001; Prinsen et al., 2007).

In addition, we found that compared with control groups using text and 2D anatomy models from the textbook, the students in both experimental groups had higher JVA ratios and better knowledge retention by interacting with 3D models on the tablet screen or AR system. Specifically, in teams with the screen-based AR, students could easily collaborate and locate specific muscles with high accuracy projected on top of their own bodies. This outcome was also highlighted in a recent meta-analysis study as collaborative learning being the most beneficial approach in any AR interventions (Garzón et al., 2020). Our study also provides further evidence that 3D visualisation technologies increase students' engagement and improve their knowledge retention in human anatomy learning (Hackett & Proctor, 2018; Luursema et al., 2008; Nicholson et al., 2006; Yammine & Violato, 2015).

Conclusion

In this paper, we introduced an automated team assessment tool based on gaze points and JVA information extracted by computer vision solutions. The results from a pilot study indicate that experimental teams who interacted with 3D digital learning tools had a high frequency of JVA and better knowledge retention outcomes than those in the control group. We also investigated the association of user study gender composition effects on JVA ratios and team test scores. We found no significant difference for JVA ratios or post-test scores among different teams with varied gender compositions.

This work was a preliminary study to automatically assess team collaboration with computer vision techniques. Like any other project, there is room for improvement. The focus was to understand collaboration on co-located situations, and shared gaze features were identified during the post-analysis process. There was a subset of collected data; since the main objective of this paper was to first understand the dyadic interactions of students, we excluded larger teams. In further planned research, a data set with team sizes larger than two would better illustrate our idea and validate our findings. In future work, it will be ideal to evaluate the effectiveness of adopting our method to other collaborative learning scenarios. Furthermore, we plan to work on multiple computer vision techniques by combining multiple image-based features, such as facial expression recognition, emotion recognition and head and body pose estimation, along with joint attention estimation to more comprehensively interpret collaboration dynamics. The findings from this work have implications in educational technology and collaborative computing by offering a novel assessment tool for team collaborations based on gaze information.

Acknowledgements

Statement on open data, ethics and conflict of interest

The study was approved by the Institutional Review Board (Protocol # HIRB00005021). Since participant information was identifiable from the image data, we were not able to share participants' data from this study based on the ethical requirements. The authors declare no conflicts of interest.

References

Andersson, J. (2001). Net effect of memory collaboration: How is collaboration affected by factors such as friendship, gender and age? *Scandinavian Journal of Psychology, 42*(4), 367–375. https://doi.org/10.1111/1467-9450.00248

Baker, R. S., & Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. *Journal of Educational Data Mining, 1*(1), 3–17. https://doi.org/10.5281/zenodo.3554657

Barmaki, R., Yu, K., Pearlman, R., Shingles, R., Bork, F., Osgood, G. M., & Navab, N. (2019). Enhancement of anatomical education using augmented reality: An empirical study of body painting. *Anatomical Sciences Education, 16*(6), 599–609. https://doi.org/10.1002/ase.1858

Barzansky, B. (1997). Educational programs in US medical schools, 1996-1997. *JAMA: The Journal of the American Medical Association, 278*(9), 744–749. https://doi.org/10.1001/jama.1997.03550090068035

Bear, J. B., & Woolley, A. W. (2011). The role of gender in team collaboration and performance. *Interdisciplinary Science Reviews, 36*(2), 146–153. https://doi.org/10.1179/030801811X13013181961473

Bente, G., Eschenburg, F., & Krämer, N. C. (2007). Virtual gaze. A pilot study on the effects of computer simulated gaze in avatar-based conversations. In R. Shumaker (Ed.), *Lecture notes in computer science: Vol. 4563. Virtual reality* (pp. 185–194). Springer. https://doi.org/10.1007/978-3-540-73335-5_21

Bertsimas, D., & Gupta, S. (2016). Fairness and collaboration in network air traffic flow management: An optimization approach. *Transportation Science*, *50*(1), 57–76. https://doi.org/10.1287/trsc.2014.0567

Brennan, S. E., Chen, X., Dickinson, C. A., Neider, M. B., & Zelinsky, G. J. (2008). Coordinating cognition: The costs and benefits of shared gaze during collaborative search. *Cognition, 106*(3), 1465–1477. https://doi.org/10.1016/j.cognition.2007.05.012

Brooks, M. (2017, December 19). More women than men enrolled in US med schools for first time. *Medscape*. https://www.medscape.com/viewarticle/890321

Bruinsma, Y., Koegel, R. L., & Koegel, L. K. (2004). Joint attention and children with autism: A review of the literature. *Mental Retardation and Developmental Disabilities Research Reviews*, *10*(3), 169–175. https://doi.org/10.1002/mrdd.20036

Bryant, T., Radu, I., & Schneider, B. (2019). A qualitative analysis of joint visual attention and collaboration with high-and low-achieving groups in computer-mediated learning. In K. Lund, G. P. Nicollai, E. Lavoué, C. Hmelo-Silver, G. Gweon, & M. Baker (Eds.), *Proceedings of the 13th International Conference on Computer Supported Collaborative Learning* (Vol. 2, pp. 923–924). International Society of the Learning Sciences. https://repository.isls.org/bitstream/1/1731/1/923-924.pdf

Carpenter, M., & Tomasello, M. (1995). Joint attention and imitative learning in children, chimpanzees, and enculturated chimpanzees. *Social Development*, *4*(3), 217–237. https://doi.org/10.1111/j.1467-9507.1995.tb00063.x

de Freitas, S., & Griffiths, M. (2007). Online gaming as an educational tool in learning and training. *British Journal of Educational Technology, 38*(3), 535–537. https://doi.org/10.1111/j.1467-8535.2007.00720.x

De Paola, M., Gioia, F., & Scoppa, V. (2018). *Teamwork, leadership and gender* (IZA Discussion Papers, No. 11861). Institute of Labor Economics. https://www.econstor.eu/bitstream/10419/185321/1/dp11861.pdf

Eagly, A. H., & Carli, L. L. (2003). The female leadership advantage: An evaluation of the evidence. *The Leadership Quarterly, 14*(6), 807–834. https://doi.org/10.1016/j.leaqua.2003.09.004

Fernandez-Sanz, L., & Misra, S. (2012). Analysis of cultural and gender influences on teamwork performance for software requirements analysis in multinational environments. *IET Software, 6*(3), 167–175. https://doi.org/10.1049/iet-sen.2011.0070

Flor, M., Yoon, S.-Y., Hao, J., Liu, L., & von Davier, A. (2016). Automated classification of collaborative problem solving interactions in simulated science tasks. In J. Tetreault, J. Burstein, C. Leacock, & H. Yannakoudakis (Eds.), *Proceedings of the 11th Workshop on Innovative Use of NLP for Building Educational Applications* (pp. 31–41). Association for Computational Linguistics. https://doi.org/10.18653/v1/W16-0504

Garzón, J., Baldiris, S., Gutiérrez, J., & Pavón, J. (2020). How do pedagogical approaches affect the impact of augmented reality on education? A meta-analysis and research synthesis. *Educational Research Review, 31,* Article 100334. https://doi.org/10.1016/j.edurev.2020.100334

Gelderblom, H., Matthee, M., Hattingh, M., & Weilbach, L. (2019). High school learners' continuance intention to use electronic textbooks: A usability study. *Education and Information Technologies, 24*(2), 1753–1776. https://doi.org/10.1007/s10639-018-9850-z

Golenhofen, N., Heindl, F., Grab-Kroll, C., Messerer, D. A., Böckers, T. M., & Böckers, A. (2020). The use of a mobile learning tool by medical students in undergraduate anatomy and its effects on assessment outcomes. *Anatomical Sciences Education, 13*(1), 8–18. https://doi.org/10.1002/ase.1878

Guo, Z., & Barmaki, R. (2019). Collaboration analysis using object detection. In C. F. Lynch, A. Merceron, M. Desmarais, & R. Nkambou (Eds.), *Proceedings of the 12th International Conference on Educational Data Mining* (pp. 695–698). Educational Data Mining. https://drive.google.com/file/d/1yznkJQl-bkP1y5sIjRjm8RwaInXNqp-k/view

Hackett, M., & Proctor, M. (2018). The effect of autostereoscopic holograms on anatomical knowledge: A randomised trial. *Medical Education, 52*(11), 1147–1155. https://doi.org/10.1111/medu.13729

Hansen, D. W., & Ji, Q. (2009). In the eye of the beholder: A survey of models for eyes and gaze. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 32*(3), 478–500. https://doi.org/10.1109/TPAMI.2009.30

Harari, D., Tenenbaum, J. B., & Ullman, S. (2018, April 30). Discovery and usage of joint attention in images. *ArXiv Preprint ArXiv:1804.04604*. Cornell University. https://arxiv.org/abs/1804.04604

Herring, S. C. (Ed.). (1996). *Computer-mediated communication: Linguistic, social, and cross-cultural perspectives*. John Benjamins Publishing. https://doi.org/10.1075/pbns.39

Hirotani, M., Stets, M., Striano, T., & Friederici, A. D. (2009). Joint attention helps infants learn new words: Event-related potential evidence. *Neuroreport, 20*(6), 600–605. https://doi.org/10.1097/WNR.0b013e32832a0a7c

Huang, K., Bryant, T., & Schneider, B. (2019). Identifying collaborative learning States using unsupervised machine learning on eye-tracking, physiological and motion sensor data. In C. F. Lynch, A. Merceron, M. Desmarais, & R. Nkambou (Eds.), *Proceedings of the 12th International Conference on Educational Data Mining* (pp. 318–323). Educational Data Mining. https://drive.google.com/file/d/1i_Ga8pmL_3R8aQOQBYKc6R4kvKN4FUMC/view

Innes, J. E., & Booher, D. E. (2016). Collaborative rationality as a strategy for working with wicked problems. *Landscape and Urban Planning, 154*, 8–10. https://doi.org/10.1016/j.landurbplan.2016.03.016

Kim, Y., D'Angelo, C., Cafaro, F., Ochoa, X., Espino, D., Kline, A., Hamilton, E., Lee, S., Butail, S., Liu, L., Trajkova, M., Tscholl, M., Hwang, J., Lee, S., & Kwon, K. (2020). Multimodal data analytics for assessing collaborative interactions. In M. Gresalfi, & I. S. Horn (Eds.), *Proceedings of the 14th International Conference on Learning Sciences* (Vol. 5, pp. 2547–2554). International Society of the Learning Sciences. https://repository.isls.org/bitstream/1/6619/1/2547-2554.pdf

Kotsiantis, S. B. (2012). Use of machine learning techniques for educational proposes: A decision support system for forecasting students' grades. *Artificial Intelligence Review, 37*(4), 331–344. https://doi.org/10.1007/s10462-011-9234-x

Lemos, R. R., Rudolph, C. M., Batista, A. V., Conceição, K. R., Pereira, P. F., Bueno, B. S., Fiuza, P. J., & Mansur, S. S. (2019). Design of a Web3D serious game for human anatomy education: A Web3D game for human anatomy education. In A. L. Krassmann, É. Amaral, F. B. Nunes, G. B. Voss, & M. C. Zunguze (Eds.), *Handbook of research on immersive digital games in educational environments* (pp. 586–611). IGI Global. https://doi.org/10.4018/978-1-5225-5790-6.ch020

Li, Q., Lau, R. W. H., Shih, T. K., & Li, F. W. B. (2008). Technology supports for distributed and collaborative learning over the internet. *ACM Transactions on Internet Technology, 8*(2), 1–24. https://doi.org/10.1145/1323651.1323656

Lian, D., Yu, Z., & Gao, S. (2018). Believe it or not, we know what you are looking at! In C. Jawahar, H. Li, G. Mori, & K. Schindler (Eds.), *Lecture notes in computer science: Vol. 11363. Asian conference on computer vision* (pp. 35–50). Springer. https://doi.org/10.1007/978-3-030-20893-6_3

Lipponen, L. (1999). The challenges for computer supported collaborative learning in elementary and secondary level: Finnish perspectives. In C. M. Hoadley & J. Roschelle (Eds.), *Proceedings of the 1999 Conference on Computer Support for Collaborative Learning* (pp. 519–528). Association for Computing Machinery. https://dl.acm.org/doi/10.5555/1150240.1150286

Luursema, J.-M., Verwey, W. B., Kommers, P. A., & Annema, J.-H. (2008). The role of stereopsis in virtual anatomical learning. *Interacting with Computers, 20*(4–5), 455–460. https://doi.org/10.1016/j.intcom.2008.04.003

Luursema, J.-M., Verwey, W. B., Kommers, P. A., Geelkerken, R. H., & Vos, H. J. (2006). Optimizing conditions for computer-assisted anatomical learning. *Interacting with Computers, 18*(5), 1123–1138. https://doi.org/10.1016/j.intcom.2006.01.005

Maresky, H., Oikonomou, A., Ali, I., Ditkofsky, N., Pakkal, M., & Ballyk, B. (2019). Virtual reality and cardiac anatomy: Exploring immersive three-dimensional cardiac imaging, a pilot study in undergraduate medical anatomy education. *Clinical Anatomy, 32*(2), 238–243. https://doi.org/10.1002/ca.23292

Marieb, E. N. (2015). *Essentials of human anatomy & physiology laboratory manual* (6th ed.). Pearson.

Marín-Jiménez, M. J., Zisserman, A., Eichner, M., & Ferrari, V. (2014). Detecting people looking at each other in videos. *International Journal of Computer Vision, 106*(3), 282–296. https://doi.org/10.1007/s11263-013-0655-7

Markus, J., Mundy, P., Morales, M., Delgado, C. E., & Yale, M. (2000). Individual differences in infant skills as predictors of child-caregiver joint attention and language. *Social Development, 9*(3), 302–315. https://doi.org/10.1111/1467-9507.00127

Meadows, L. A., Sekaquaptewa, D., & Paretti, M. C. (2015). Interactive panel: Improving the experiences of marginalized students on engineering design teams. In B. M. Holloway (Ed.), *Proceedings of the ASEE Annual Conference & Exposition* (Vol. 26, pp. 1–23). American Society for Engineering Education. https://doi.org/10.18260/p.24344

Meiksins, P., Beddoes, K., Layne, P., McCusker, M., & Camargo, E. (2015, November 30). Women in engineering: A review of the 2014 literature. *Consulting – Specifying Engineer*. https://www.csemag.com/articles/women-in-engineering-a-review-of-the-2014-literature/

Miura, G., & Okada, S. (2019). Task-independent multimodal prediction of group performance based on product dimensions. In W. Gao, M. Meng, M. Turk, S. R. Fussell, B. Schuller, Y. Song, & K. Yu (Eds.), *Proceedings of the 2019 International Conference on Multimodal Interaction* (pp. 264–273). Association for Computing Machinery. https://doi.org/10.1145/3340555.3353729

Morris, S. B. (2008). Estimating effect sizes from pretest-posttest-control group designs. *Organizational Research Methods*, *11*(2), 364–386. https://doi.org/10.1177/1094428106291059

Mukherjee, S. S., & Robertson, N. M. (2015). Deep head pose: Gaze-direction estimation in multimodal video. *IEEE Transactions on Multimedia*, *17*(11), 2094–2107. https://doi.org/10.1109/TMM.2015.2482819

Murray, G., & Oertel, C. (2018). Predicting group performance in task-based interaction. In S. K. D'mello, P. Georgiou, S. Scherer, E. M. Provost, M. Soleymani, & M. Worsley (Eds.), *Proceedings of the 20th ACM International Conference on Multimodal Interaction* (pp. 14–20). Association for Computing Machinery. https://doi.org/10.1145/3242969.3243027

Ng, J., Hu, X., Luo, M., & Chu, S. K. W. (2019). Relations among participation, fairness and performance in collaborative learning with Wiki-based analytics. *Proceedings of the Association for Information Science and Technology, 56*(1), 463–467. https://doi.org/10.1002/pra2.48

Nicholson, D. T., Chalk, C., Funnell, W. R. J., & Daniel, S. J. (2006). Can virtual reality improve anatomy education? A randomised controlled study of a computer-generated three-dimensional anatomical ear model. *Medical Education, 40*(11), 1081–1087. https://doi.org/10.1111/j.1365-2929.2006.02611.x

Ochoa, X., Chiluiza, K., Méndez, G., Luzardo, G., Guamán, B., & Castells, J. (2013). Expertise estimation based on simple multimodal features. In J. Epps, F. Chen, S. Oviatt, K. Mase, A. Sears, K. Jokinen, & B. Schuller (Eds.), *Proceedings of the 15th ACM on International Conference on Multimodal Interaction* (pp. 583–590). Association for Computing Machinery. https://doi.org/10.1145/2522848.2533789

Okita, S., Bailenson, J., & Schwartz, D. (2008). Mere belief of social action improves complex learning. In P. A. Kirschner, J. J. van Merriënboer, & T. de Jong (Eds.), *Proceedings of the 8th International Conference for The Learning Sciences* (Vol. 2, pp. 132–139). International Society of the Learning Sciences (ISLS). https://repository.isls.org/bitstream/1/3144/1/132-139.pdf

Otsuka, K., Kasuga, K., & Köhler, M. (2018). Estimating visual focus of attention in multiparty meetings using deep convolutional neural networks. In S. K. D'mello, P. Georgiou, S. Scherer, E. M. Provost, M. Soleymani, & M. Worsley (Eds.), *Proceedings of the 20th ACM International Conference on Multimodal Interaction* (pp. 191–199). Association for Computing Machinery. https://doi.org/10.1145/3242969.3242973

Patacchiola, M., & Cangelosi, A. (2017). Head pose estimation in the wild using convolutional neural networks and adaptive gradient methods. *Pattern Recognition*, *71*, 132–143. https://doi.org/10.1016/j.patcog.2017.06.009

Pietinen, S., Bednarik, R., & Tukiainen, M. (2010, May). Shared visual attention in collaborative programming: A descriptive analysis. In Y. Dittrich, C. de Souza, M. Korpela, H. Sharp, J. Singer, & H. W. Theophilus (Eds), *Proceedings of the 2010 ICSE Workshop on Cooperative and Human Aspects of Software Engineering* (pp. 21–24). Association for Computing Machinery. https://doi.org/10.1145/1833310.1833314

Prinsen, F. R., Volman, M. L., & Terwel, J. (2007). Gender-related differences in computer-mediated communication and computer-supported collaborative learning. *Journal of Computer Assisted Learning*, *23*(5), 393–409. https://doi.org/10.1111/j.1365-2729.2007.00224.x

Rabbitt, P., Donlan, C., Watson, P., McInnes, L., & Bent, N. (1995). Unique and interactive effects of depression, age, socioeconomic advantage, and gender on cognitive performance of normal healthy older people. *Psychology and Aging, 10*(3), 307–313. https://doi.org/10.1037/0882-7974.10.3.307

Recasens, A., Khosla, A., Vondrick, C., & Torralba, A. (2015). Where are they looking? In C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems* (pp. 199–207). Neural Information Processing Systems Foundation. http://papers.nips.cc/paper/5848-where-are-they-looking

Rezende, E., Ruppert, G., Carvalho, T., Ramos, F., & De Geus, P. (2017). Malicious software classification using transfer learning of resnet-50 deep neural network. In X. Chen (Ed.), *Proceedings of the 16th IEEE International Conference on Machine Learning and Applications* (pp. 1011–1014). IEEE. https://doi.org/10.1109/ICMLA.2017.00-19

Rosé, C., Wang, Y.-C., Cui, Y., Arguello, J., Stegmann, K., Weinberger, A., & Fischer, F. (2008). Analyzing collaborative learning processes automatically: Exploiting the advances of computational linguistics in computer-supported collaborative learning. *International Journal of Computer-Supported Collaborative Learning, 3*(3), 237–271. https://doi.org/10.1007/s11412-007-9034-0

Santini, T., Fuhl, W., & Kasneci, E. (2017, May). CalibMe: Fast and unsupervised eye tracker calibration for gaze-based pervasive human-computer interaction. In G. S. Mark, C. Fussell, Lampe, M. C. Schraefel, J. P. Hourcade, C. Appert, & D. Wigdor (Eds.), *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (pp. 2594–2605). Association for Computing Machinery. https://doi.org/10.1145/3025453.3025950

Schaf, F. M., Müller, D., Bruns, F. W., Pereira, C. E., & Erbe, H.-H. (2009). Collaborative learning and engineering workspaces. *Annual Reviews in Control, 33*(2), 246–252. https://doi.org/10.1016/j.arcontrol.2009.05.002

Schaie, K. W., & Willis, S. L. (1993). Age difference patterns of psychometric intelligence in adulthood: Generalizability within and across ability domains. *Psychology and Aging, 8*(1), 44. https://doi.org/10.1037/0882-7974.8.1.44

Schneider, B., & Pea, R. (2013). Real-time mutual gaze perception enhances collaborative learning and collaboration quality. *International Journal of Computer-Supported Collaborative Learning*, *8*(4), 375–397. https://doi.org/10.1007/s11412-013-9181-4

Schneider, B., Sharma, K., Cuendet, S., Zufferey, G., Dillenbourg, P., & Pea, R. (2018). Leveraging mobile eye-trackers to capture joint visual attention in co-located collaborative learning groups. *International Journal of Computer-Supported Collaborative Learning, 13*(3), 241–261. https://doi.org/10.1007/s11412-018-9281-2

Silva, J. N., Southworth, M., Raptis, C., & Silva, J. (2018). Emerging applications of virtual reality in cardiovascular medicine. *JACC: Basic to Translational Science, 3*(3), 420–430. https://doi.org/10.1016/j.jacbts.2017.11.009

Soller, A., Lesgold, A., Linton, F., & Goodman, B. (1999). What makes peer interaction effective? Modeling effective communication in an intelligent CSCL. In S. E. Brennan, A. Giboin, & D. Traum (Eds.), *Proceedings of 1999 AAAI Fall Symposium: Psychological Models of Communication in Collaborative Systems* (pp. 116–123). AAAI Press. https://www.aaai.org/Papers/Symposia/Fall/1999/FS-99-03/FS99-03-017.pdf

Spikol, D., Ruffaldi, E., & Cukurova, M. (2017). Using multimodal learning analytics to identify aspects of collaboration in project-based learning. In B. K. Smith, M. Borge, E. Mercier, & K. Y. Lim (Eds.), *Proceedings of the 2017 Conference on Computer Support for Collaborative Learning* (pp. 263–270). International Society of the Learning Sciences. https://repository.isls.org/bitstream/1/240/1/37.pdf

Subburaj, S. K., Stewart, A. E. B., Ramesh Rao, A., & D'Mello, S. K. (2020). Multimodal, multiparty modeling of collaborative problem solving performance. In K. Truong, D. Heylen, M. Czerwinski, N. Berthouze, M. Chetouani, & M. Nakano (Eds.), *Proceedings of the 2020 International Conference on Multimodal Interaction* (pp. 423–432). Association for Computing Machinery. https://doi.org/10.1145/3382507.3418877

Sung, H.-Y., & Hwang, G.-J. (2013). A collaborative game-based learning approach to improving students' learning performance in science courses. *Computers & Education, 63*, 43–51. https://doi.org/10.1016/j.compedu.2012.11.019

Tao, C., Zhang, Q., & Zhou, Y. (2019). Collaborative learning with limited interaction: Tight bounds for distributed exploration in multi-armed bandits. In L. O'Conner (Ed.), *Proceedings of the 2019 IEEE 60th Annual Symposium on Foundations of Computer Science* (pp. 126–146). IEEE. https://doi.org/10.1109/FOCS.2019.00017

van der Meulen, H., Varsanyi, P., Westendorf, L., Kun, A. L., & Shaer, O. (2016). Towards understanding collaboration around interactive surfaces: Exploring joint visual attention. In J. Rekomoto, T. Igarashi, J. O. Wobbrock, & D. Avrahami (Eds.), *Proceedings of the 29th Annual Symposium on User Interface Software and Technology* (pp. 219–220). Association for Computing Machinery. https://doi.org/10.1145/2984751.2984778

Van Rheden, V., Maurer, B., Smit, D., Murer, M., & Tscheligi, M. (2017). LaserViz: Shared gaze in the Co-located physical world. In M. Inakage, H. Ishii, E. Y. Do, J. Steimle, O. Shaer, K. Kunze, & R. Peiris (Eds.), *Proceedings of the Eleventh International Conference on Tangible, Embedded, and Embodied Interaction* (pp. 191–196). Association for Computing Machinery. https://doi.org/10.1145/3024969.3025010

Vrzakova, H., Amon, M. J., Stewart, A. E. B., & D'Mello, S. K. (2019). Dynamics of visual attention in multiparty collaborative problem solving using multidimensional recurrence quantification analysis. In S. Brewster, G. Fitzpatrick, A. Cox, & V. Kostakos (Eds.), *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1–14). Association for Computing Machinery. https://doi.org/10.1145/3290605.3300572

Wahn, B., Schwandt, J., Krüger, M., Crafa, D., Nunnendorf, V., & König, P. (2016). Multisensory teamwork: Using a tactile or an auditory display to exchange gaze information improves performance in joint visual search. *Ergonomics, 59*(6), 781–795. https://doi.org/10.1080/00140139.2015.1099742'

Webb, N. M., Troper, J. D., & Fall, R. (1995). Constructive activity and learning in collaborative small groups. *Journal of Educational Psychology*, *87*(3)*,* 406–423. https://doi.org/10.1037/0022-0663.87.3.406

Wegge, J., Roth, C., Neubach, B., Schmidt, K.-H., & Kanfer, R. (2008). Age and gender diversity as determinants of performance and health in a public organization: The role of task complexity and group size. *Journal of Applied Psychology, 93*(6), 1301–1313. https://doi.org/10.1037/a0012680

Wendler, D. (2006). One-time general consent for research on biological samples. *BMJ: British Medical Journal, 332*(7540), 544–547. https://doi.org/10.1136/bmj.332.7540.544

Whalen, C., & Schreibman, L. (2003). Joint attention training for children with autism using behavior modification procedures. *Journal of Child Psychology and Psychiatry, 44*(3), 456–468. https://doi.org/10.1111/1469-7610.00135

Yammine, K., & Violato, C. (2015). A meta-analysis of the educational effectiveness of three-dimensional visualization technologies in teaching anatomy. *Anatomical Sciences Education, 8*(6), 525–538. https://doi.org/10.1002/ase.1510

Yücel, Z., Salah, A. A., Meriçli, Ç., Meriçli, T., Valenti, R., & Gevers, T. (2013). Joint attention by gaze interpolation and saliency. *IEEE Transactions on Cybernetics, 43*(3), 829–842. https://doi.org/10.1109/TSMCB.2012.2216979

Zhao, Q., Sheng, T., Wang, Y., Tang, Z., Chen, Y., Cai, L., & Ling, H. (2019). M2det: A single-shot object detector based on multi-level feature pyramid network. In P. Stone, P. V. Hentenryck, & Z. H. Zhou (Eds.), *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 33, pp. 9259–9266). AAAI Press. https://doi.org/10.1609/aaai.v33i01.33019259

Zhu, X., & Ramanan, D. (2012). Face detection, pose estimation, and landmark localization in the wild. In R. Chellappa, B. Kimia, S. C. Zhu, S. Belongie, A. Blake, J. Luo, & A. Yuille (Eds.), *Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 2879–2886). IEEE. https://doi.org/10.1109/CVPR.2012.6248014

Zou, K. H., Tuncali, K., & Silverman, S. G. (2003). Correlation and simple linear regression. *Radiology, 227*(3), 617–628. https://doi.org/10.1148/radiol.2273011499

Corresponding author: Roghayeh Barmaki, rlb@udel.edu