Heliyon 8 (2022) e12335

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Eye tracking metrics of orthopedic surgeons with different competency levels who practice simulation-based hip arthroscopic procedures

Bohong Cai [a,1], Na Xu [a,1], Shengfeng Duan [a], Jiahui Yi [a], Boon Huat Bay [b], Fangyuan Shen [c], Ning Hu [d], Peng Zhang [e], Jie Chen [d,**], Cheng Chen [d,*]

[a] Department of Industrial and Product Design, School of Design, Sichuan Fine Arts Institute, Chongqing, China
[b] Department of Anatomy, Yong Loo Lin School of Medicine, National University of Singapore, Singapore
[c] State Key Laboratory of Oral Diseases and National Clinical Research Center for Oral Diseases, West China Hospital of Stomatology, Sichuan University, Chengdu, China
[d] Department of Orthopedics, The First Affiliated Hospital of Chongqing Medical University, Chongqing, China
[e] Department of Orthopedics, Sichuan Province Orthopedic Hospital, Chengdu, China

ARTICLE INFO

Keywords:
Hip arthroscopy
Simulation
Eye tracking
Surgical training
Technical assessment

ABSTRACT

Objective: This study aimed to investigate the feasibility of using eye tracking data to identify orthopedic trainees' technical proficiency in hip arthroscopic procedures during simulation-based training.
Design: A cross sectional study.
Setting: A simulation-based training session for hip arthroscopy was conducted. Eye tracking devices were used to record participants' eye movements while performing simulated operations. The NASA Task Load Index survey was then used to measure subjective opinions on the perceived workload of the training. Statistical analyses were performed to determine the significance of the eye metrics and survey data.
Participants: A total of 12 arthroscopic trainees, including resident doctors, junior specialist surgeons, and consultant surgeons from the Department of Orthopedics in five hospitals, participated in this study. They were divided into three subgroups based on their prior clinical experience.
Results: Significant differences, including those for dwell time, number of fixations, number of saccades, saccade duration, peak velocity of the saccade, and pupil entropy, were observed among the three subgroups. Additionally, there were clear trends in the perceived workload of the simulation-based training based on feedback from the participants.
Conclusion: Based on this preliminary study, a correlation was identified between the eye tracking metrics and participants' experience levels. Hence, it is feasible to apply eye tracking data as a supplementary objective assessment tool to benchmark the technical proficiency of surgical trainees in hip arthroscopy, and enhance simulation-based training.

1. Introduction

Hip arthroscopic surgery is rapidly developing as a technique used in orthopedic surgery to diagnose and treat many hip-related conditions [1]. Beyond the general difficulties of arthroscopic surgery, such as reduced tactile feedback, narrowed vision of the camera, limited degrees of freedom for manipulation, and the difficulty of integrating three-dimensional (3D) information based on two-dimensional monitors [2, 3, 4], hip arthroscopic techniques are compounded by challenges involving safe access to specific anatomical sites, time constraints for surgery, and the requirements for specialized instrumentation [4]. Owing to these difficulties, orthopedic trainees and surgeons are required to spend significantly more time to practice the relevant technical skills to achieve competency.

Previous research findings have inferred that a surgeon would be required to perform more than 100 cases of hip arthroscopic surgery to

* Corresponding author.
** Corresponding author.
E-mail addresses: chenjie_doc@163.com (J. Chen), chenchengdoc@126.com (C. Chen).
[1] Equal contribution.

https://doi.org/10.1016/j.heliyon.2022.e12335
Received 17 May 2022; Received in revised form 16 October 2022; Accepted 6 December 2022

achieve optimal clinical outcomes [5, 6]. This learning curve has become smoother with developments in simulation technology. Instead of only relying on operating room experience to master the appropriate surgical skills, medical professionals can practice skills repeatedly in a safe environment [7, 8]. Various types of simulators that enhance training efficiency have been used to facilitate the training of hip arthroscopic skills [9, 10, 11], thus shortening the training duration without compromising the quality of patient care.

As acquiring clinical skills by simulation-based training becomes increasingly popular [12], objective assessment methods to assess the trainees' surgical skills are necessary to ensure that they have achieved adequate levels of competency [13]. Systematic and objective assessments can help medical educators to structure competency-based training programs [14], to benefit surgeon certification processes and improve quality of healthcare and patient safety. Facilitated by evaluation tools, such as the Objective Structured Assessment of Technical Skills [15] and Arthroscopic Surgical Skill Evaluation Tool [16] the assessment process has become more accurate and impartial. However, as most evaluation tools rely on the subjective opinions of examiners, the risk of bias cannot be completely avoided.

In recent years, eye tracking has shown its potential as a measurement tool for assessing the technical skills of surgery [17, 18, 19]. The eye tracking system records and analyses eye movements and pupil behavior during operative procedures. It has been considered a suitable assessment method as it is based on objective data, which avoids the potential bias of a subjective grading approach. Studies have revealed that eye tracking metrics show strong associations with perceived workload [20, 21]. For example, a higher fixation rate and proportion of fixation within an area of interest (AOI) indicate a greater degree of focused attention, and more elaborate visual representation during a task, which are traits associated with a more experienced operator [18, 22]. Furthermore, eye metrics, such as the overlap of gaze and fixation locations, can be considered as surrogate markers for identifying the level of expertise of surgeons [23]. Fixation metrics and scan path patterns observed in medical professionals are able to reflect their operational behaviors, which can generate benchmarks for improving training strategies [24]. Thus, tracking the eye metrics of medical trainees and educators can provide a better insight to learning behavior and assist in gauging the trainees' proficiency level in specific technical skills.

Moreover, the increasing use of simulation-based training to smoothen the steep learning curve of hip arthroscopy is appealing for evaluating eye metrics of surgeons during simulated operations. Data obtained from these investigations can be valuable in revealing intraoperative cognitive processes of the arthroscopic surgeon, especially while performing complex and challenging procedures. These studies will be beneficial to medical educators in providing a starting point to improve program design and optimizing the usage of simulation in surgical training. Therefore, the aim of this study was to identify the differences in eye metrics for surgeons with varying levels of competency when performing hip arthroscopic procedures in a simulated setting.

2. Results

A total of 12 arthroscopic surgeons, including resident doctors, junior specialist surgeons, and consultant surgeons from the Department of Orthopedics in five hospitals, were included in this study. All participants were male and right-handed. The novice group performed lesser arthroscopic operations. Only surgeons with specific experience in hip arthroscopy were classified in the experienced group, or else, they were placed in the intermediate group (details as shown in Table 1).

All participants performed the simulated operation on a 3D printed simulator over 3 days. Eye movements were recorded simultaneously. Varied features of eye movements were recorded on different AOIs. As four participants, one from the intermediate group and three from the experienced group, did not use fluoroscopy to assist portal placement, the experienced group was not included in the analysis of fluoroscopy-related metrics.

Table 1. Clinical experience of subgroups.[a]

	Novice group	Intermediate group	Experienced group
n	3	5	4
Performed arthroscopic surgery for hip (mean ± SD)	0	0	22.5 ± 5.00
Performed arthroscopic surgery for shoulder (mean ± SD)	16.67 ± 20.82	82 ± 73.28	105 ± 42.03
Performed arthroscopic surgery for knee (mean ± SD)	46.67 ± 35.12	190 ± 174.64	128.33 ± 122.98

[a] Data are presented as the mean number of arthroscopic procedures performed.

2.1. Eye tracking metrics

The novice group had the longest dwell time, followed by the intermediate and experienced groups, at the operative site and arthroscopic field (Table 2). For the fluoroscopic field, the novice group had a significantly longer dwell time than the intermediate group (Figure 1A; median, 125.206 [IQR: 26.5–138.4] vs. median, 12.252 [IQR: 4.9–20.7]; $P = 0.034$, respectively). The results demonstrated a clear trend that the groups with more prior surgical experience had shorter dwell times during the simulated operation.

The number of fixations for the operative site significantly differed between the three subgroups (Table 3). The novice and intermediate groups showed significant difference in fixations for the fluoroscopic field (Figure 1B; median, 56 [IQR: 34–89] vs. median, 10.5 [IQR: 6.3–14.8]; $P = 0.034$, respectively). In general, the number of fixations was reduced with increased surgical experience; only the intermediate group showed a higher number of fixations over the novice group for the arthroscopic field.

The number of saccades was significantly different for the fluoroscopic field, as the novice group had more saccades than the intermediate group (Figure 1C; median, 30 [IQR: 19–44] vs. median, 9 [IQR: 3.5–16.0]; $P = 0.034$, respectively). For the operative site, the number of saccades was decreased with increased surgical experience, although this was not statistically significant (Table 4). However, for the arthroscopic field, the novice group had fewer saccades than the intermediate and experienced groups.

The saccade duration showed significant differences between the three groups. Participants with more surgical experience had a shorter saccade duration ($P < 0.001$), as shown in Table 5. Notably, the experienced group had a shorter saccade duration than the intermediate group in the operative site, and shorter duration than the novice group in the arthroscopic field. In the fluoroscopic field, no significant difference were observed between the novice and intermediate groups (Figure 1D; median, 20 [IQR: 20–60] vs. median, 20 [IQR: 20–40]; P = n.s., respectively).

Similarly, significant differences were observed in the peak velocity of saccades between the three groups (Table 6). In general, peak velocity values for the operative site and arthroscopic field increased with the participants' clinical experience ($P < 0.001$); only the intermediate group had a slightly lower value than the novice group in the arthroscopic field. A similar difference was observed between the novice and intermediate groups for the fluoroscopic field (Figure 1E; median, 177.06 [IQR: 122.85–274.29] vs. median, 193.68 [IQR: 134.49–292.66]; $P = 0.048$, respectively).

Pupil entropy was observed to be significantly different for the operative site and arthroscopic field; participants with more surgical experience had a smaller pupil entropy (Table 7). For the fluoroscopic

Table 2. Details on the dwell time (second) for three subgroups.

Areas of interest	Novice Group		Intermediate Group		Experienced Group		*P* value
	Median	IQR	Median	IQR	Median	IQR	
Operative Site	1114.180	617.3–1823.7	911.355	312.2–1012.7	465.957	283.6–637.3	0.046
Arthroscopic Field	1845.896	1150.8–1866.5	1148.459	835.6–1176.6	578.972	500.7–732.7	0.002

IQR: interquartile range.

field, no statistical significance was found between the novice and intermediate groups (Figure 1F; median, 0.047 [IQR: 0.0239–0.1110] vs. median, 0.001 [IQR: 0.0003–0.0016]; P = n.s., respectively).

For pairwise comparisons, the dwell time for the arthroscopic field showed significant differences between the novice and experienced groups ($P = 0.006$), and the experienced group had a shorter dwell time. At the operative site, the saccade duration was significantly different when comparison was made between any two of the three groups ($P < 0.05$), and the peak velocity of saccade was significantly different in the experienced group compared to both the intermediate and novice groups ($P < 0.001$). Similarly, pairwise comparisons of the peak velocity of saccade and saccade duration in the arthroscopic field showed significant differences when comparing any two of the three groups ($P < 0.001$).

2.2. Cognitive load

Based on the NASA-TLX survey, no significant differences were discovered between the three subgroups (Table 8). However, clear trends were observed in most dimensions. Participants with less experience had higher levels of mental effort and frustration. The experienced group rated physical and temporal demand lower than groups with less experience.

3. Discussion

Eye metrics have been previously demonstrated to correlate with participants' cognitive load when performing simulated operations [21]. This preliminary study, which compared eye tracking data between groups with varying clinical experience in a simulation-based training, established specific correlations that denoted the participants' experience level and eye metrics data, which could be used as an indicator for determining participants' proficiency in hip arthroscopic surgery.

Participants with more surgical experience had a shorter duration and fewer number of fixations at the operative site. Participants presented with longer dwell times and more fixations in response to a more challenging task [25]. An association between the dwell time and information processing load has been previously reported [17, 26], where an increased dwell time and workload were correlated. Compared to the experienced group, the novice and intermediate groups had less clinical experience in arthroscopic operations, particularly, hip arthroscopy. Participants in the latter two groups were generally unfamiliar with the techniques required for the simulated training task. Therefore, it was unsurprising that they required more cognitive load when performing the assigned tasks, which correlated with the finding that the simulated operation was more demanding for less experienced participants.

Similar to the fixation metrics, the number of saccades observed was an important indicator of workload, as an increased number of saccades may suggest an increasing mental workload for the task [27]. In this study, less skilled participants had a higher number of saccades for the operative site, indicating that in these two AOIs, less skilled participants experienced increased mental cognitive load when performing the task. Notably, we observed the opposite phenomenon for the arthroscopic field, where the novice group had the lowest number of saccades between the three subgroups. This could be explained by a previous study that showed that an increased mental workload did not always equate to an increased number of saccades [28], as inexperienced participants may require more time to perceive information (which reduces the number of saccades).

Based on the precept established in previous studies [29, 30, 31, 32] viewing behavior can be divided into two categories: early and later viewing. Early viewing is characterised by more saccades and a shorter duration of fixation; as much information as possible is collected during the early phase. In contrast, later viewing has fewer saccades and a longer fixation duration to process more in-depth and object-centered analyses [32]. In this study, the novice group had the fewest saccades and longest fixation duration for the arthroscopic field, which correlated with a later viewing pattern. This observation suggested that novices were required to expend more work analyzing information from the arthroscopic screen (which was the main tool for accomplishing intra-joint operations), to complete the simulated task, compared to the intermediate and experienced groups.

In this study, other saccadic-related parameters, such as saccade duration and peak velocity of saccade, have proven that trainees with less surgical experience perceived an increased workload than more experienced trainees. Generally, less skilled trainees had a longer saccade duration and lower peak velocity of saccade than more experienced trainees; these findings were consistent with earlier studies [33, 34, 35], which reflected the status of attention span and working memory of the trainees when performing the tasks. Compared to more experienced trainees, the difficulty level of the simulated task was higher for novice trainees, which resulted in increasing attentiveness and working memory. In contrast, experienced surgeons distributed their attention more effectively by simultaneously observing the necessary locations at the appropriate time, to reduce memory impairment.

In the AOIs for the operative site and arthroscopic field, pupil entropy was discovered to be significantly higher in less experienced groups, which was consistent with earlier findings reported in surgeons who performed inguinal hernia repair [36]. Compared to surgeons from the experienced group, less experienced participants required more information processing during the task; thus, rendering pupillary changes less predictable, which meant that less experienced participants were normally associated with higher cognitive load when performing tasks [37]. The findings on pupil entropy were similar to other examined eye metrics, where participants with increasing experience levels required less cognitive load to perform simulated hip arthroscopic procedures.

Despite no observed significant differences in the cognitive load regarding mental demands, effort, and frustration, experienced participants had lower scores. Less experienced participants with inadequate experience in hip arthroscopy were more challenged when performing simulated procedures, which resulted in a higher mental demand, significantly more effort required to accomplish the task, and frequent frustration. However, given more practice, trainees could develop intraoperative mentally embedded movement coordination and cognitive load reduction [38]. Although the scores for physical and temporal demands did not decrease with improved experience levels, the experienced group had a lower score than the other two groups, which may indicate that the experience level of the orthopedic surgeon was important during arthroscopic simulations.

Eye tracking metrics can be used as an additional objective determinant of orthopedic trainees' technical proficiency and quality of training for arthroscopic procedures. With the wide-spread use of simulation-based arthroscopy, eye tracking data could be used to assess the learning efficiency of individuals using simulation-based training for

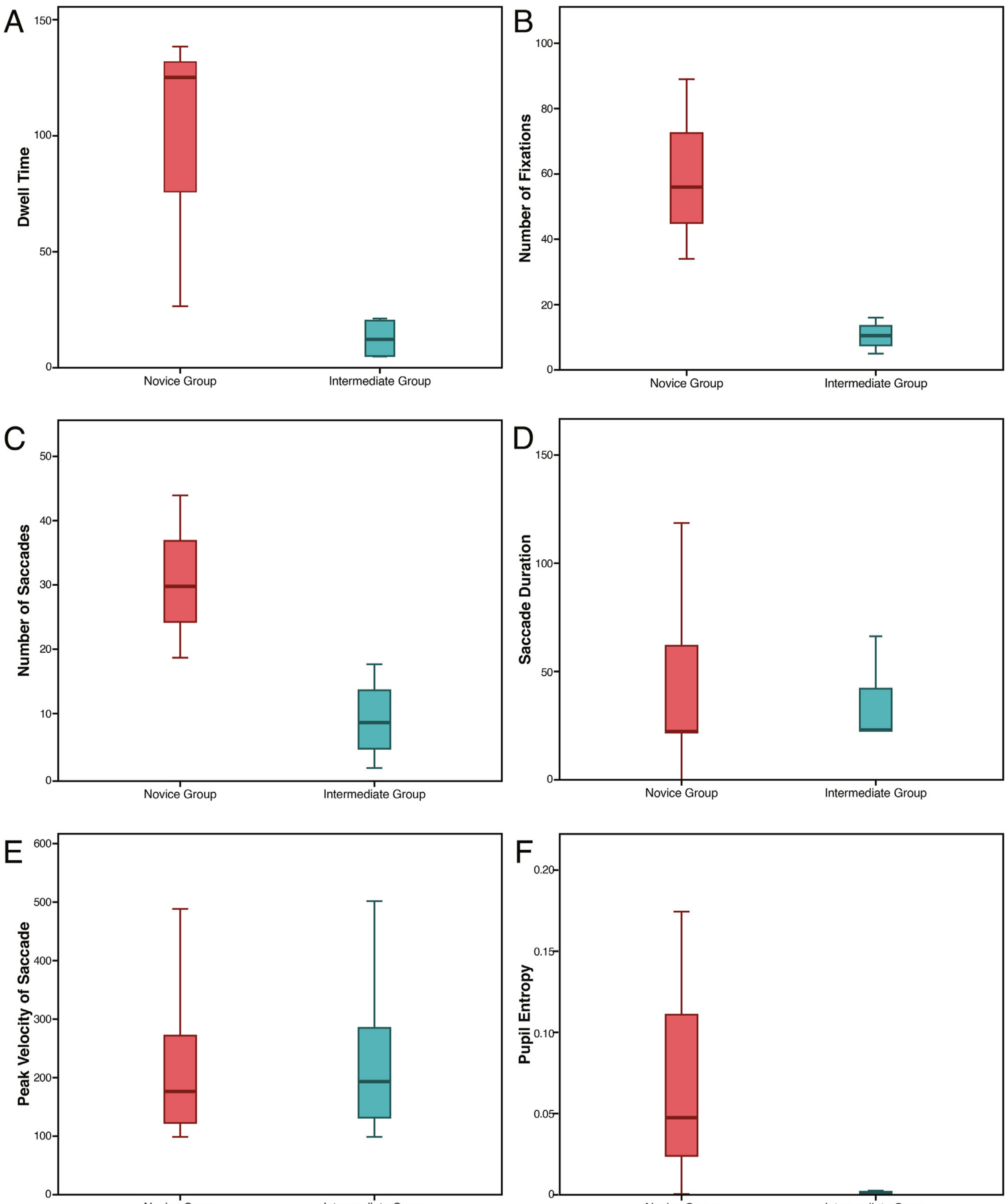

Figure 1. Box plots for eye tracking metrics for the fluoroscopic field of the novice and intermediate group. (A) Box plot of the dwell time; (B) Box plot of the number of fixations; (C) Box plot of the number of saccades; (D) Box plot of the saccade duration; (E) Box plot of the peak velocity of saccade; (F) Box plot of the pupil entropy. The box denotes the spread of data collated through their quartiles and the whisker indicate variability of the data outside the upper and lower quartiles.

honing their surgical skills. Moreover, by comparing eye tracking data of actual surgeries with simulation-based training, medical educators could evaluate the quality of the training program and the simulator used. Eye tracking metrics can be an objective reference for informing educators and researchers of the effectiveness of simulation-based training, generating similar physiological feedback to trainees, compared to an actual operation.

3.1. Limitations

This study had several limitations. First, the number of participants was small, which may have resulted in a Type II error for some variables (e.g., temporal demand on NASA-TLX). Thus, increasing the sample size would enhance the value of the study. Second, some participants (one surgeon from the intermediate group and three from the

Table 3. Details on the number of fixations (count) for three subgroups.

Areas of interest	Novice Group		Intermediate Group		Experienced Group		*P* value
	Median	IQR	Median	IQR	Median	IQR	
Operative Site	629.0	434.0–809.0	476.0	222.0–561.0	356.5	258.3–407.5	0.046
Arthroscopic Field	366.0	301.0–676.0	450.0	358.0–684.5	382.0	278.5–465.3	0.712

IQR: interquartile range.

Table 4. Details on the number of saccades (count) for three subgroups.

Areas of interest	Novice Group		Intermediate Group		Experienced Group		*P* value
	Median	IQR	Median	IQR	Median	IQR	
Operative Site	430.0	346.0–707.0	422.0	138.5–528.0	302.5	223.3–356.3	0.121
Arthroscopic Field	132.0	107.0–274.0	353.0	204.5–551.0	334.0	228.0–411.5	0.210

IQR: interquartile range.

Table 5. Details of saccade duration (millisecond) for three subgroups.

Areas of interest	Novice Group		Intermediate Group		Experienced Group		*P* value
	Median	IQR	Median	IQR	Median	IQR	
Operative site	40	20–60	40	20–60	20	20–40	<0.001
Arthroscopic field	40	20–60	20	20–40	20	20–20	<0.001

IQR: interquartile range.

Table 6. Details of peak velocity of saccade (degree per second) for three subgroups.

Areas of interest	Novice Group		Intermediate Group		Experienced Group		*P* value
	Median	IQR	Median	IQR	Median	IQR	
Operative site	191.01	132.9–292.2	198.42	136.7–299.4	224.19	148.1–314.0	<0.001
Arthroscopic field	252.49	152.9–349.9	213.27	139.5–321.7	288.61	198.7–367.9	<0.001

IQR: interquartile range.

Table 7. Details of pupil entropy (bit) for three subgroups.

Areas of interest	Novice Group		Intermediate Group		Experienced Group		*P* value
	Median	IQR	Median	IQR	Median	IQR	
Operative site	0.124	0.118–0.175	0.013	0.003–0.154	0.002	0.001–0.003	0.015
Arthroscopic field	0.129	0.086–0.148	0.018	0.007–0.038	0.003	0.002–0.006	0.010

IQR: interquartile range.

Table 8. Details on the NASA-TLX (score) for three subgroups.

Subscales	Novice Group		Intermediate Group		Experienced Group		*P* value
	Median	IQR	Median	IQR	Median	IQR	
Mental demand	7.5	6.0–9.8	7.0	5.5–7.5	5.0	3.0–7.0	0.317
Physical demand	6.0	2.8–8.5	6.0	5.5–6.5	3.0	2.3–5.3	0.324
Temporal demand	6.0	5.0–8.5	5. 0	5.0–7.0	4.0	2.3–5.0	0.083
Performance	8.5	6.3–10.0	6.0	5.5–8.0	7.0	6.0–9.5	0.939
Effort	8.0	5.5–9.8	7.0	6.5–7.5	6.0	4.5–6.8	0.164
Frustration	7.5	5.5–9.5	7.0	5.5–8.5	5.0	2.0–8.0	0.227

IQR: interquartile range.

experienced group) did not use fluoroscopy in their simulated practice, which further reduced the available data set for statistical analyses for the fluoroscopic field, and may have affected the accuracy of the results. In future, improving the study design and applying a stricter study protocol would be helpful in mitigating this issue. Third, in contrast to actual surgeries, the simulator used in this study was a task-specific trainer, which is used specifically for some important steps in hip arthroscopy; other steps were excluded. As the simulation practice

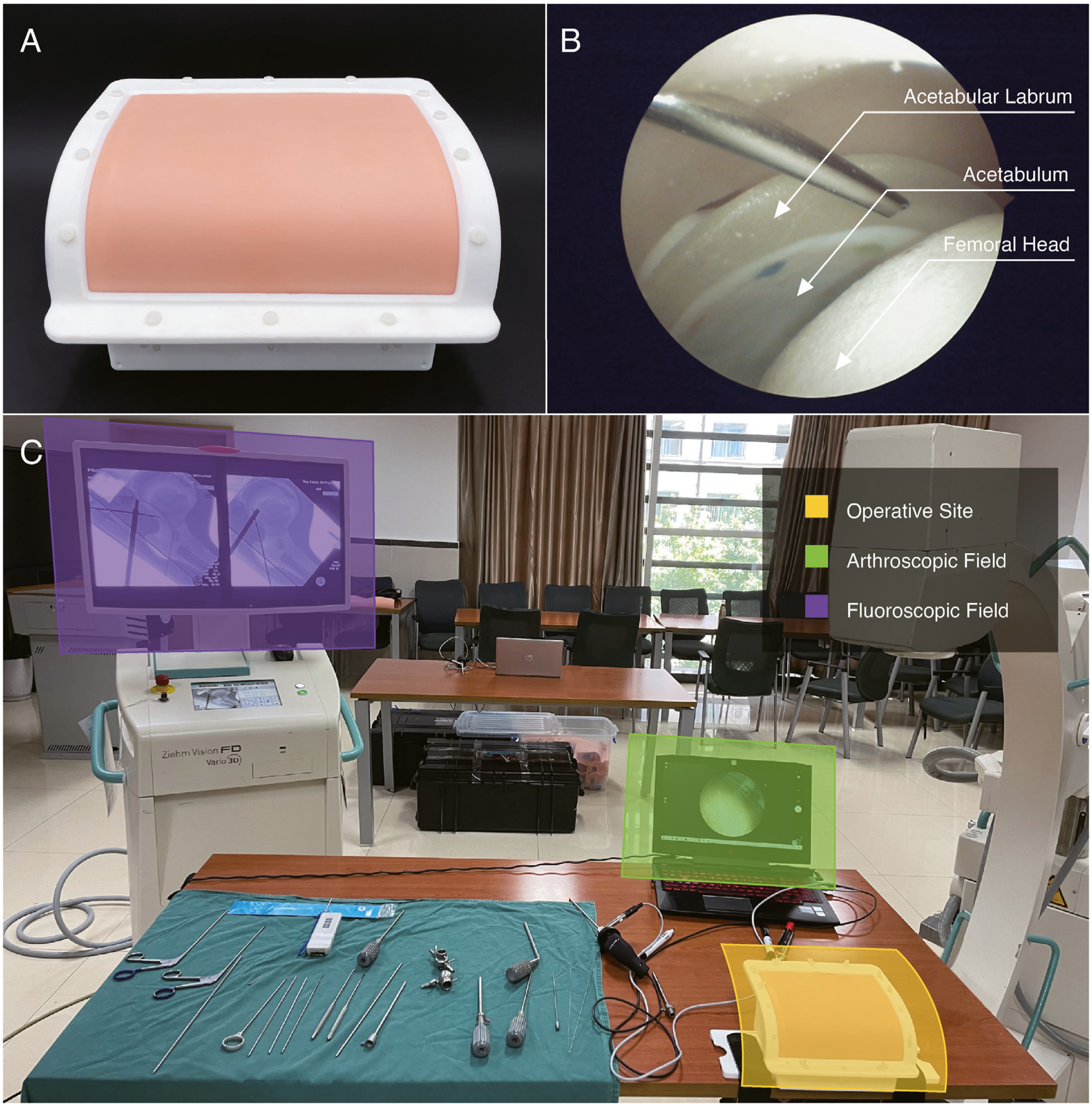

Figure 2. The equipments and settings for this study. (A) Assembled simulator with a silicon component overlaid on the bony structures; (B) Arthroscopic view showing the intra-articular space of the simulator; (C) Demonstration of areas of interest for the simulated operation.

Table 9. Measurement metrics of eye movement data.

Metrics	Description	Unit
Dwell time	Sum of durations from all fixations and saccades when visualizing the AOI	Second
Number of fixations	Total number of times when gazing at the AOI	Count
Number of saccades	Total number of saccades counted when visualizing the AOI	Count
Saccade duration	The time taken to complete the saccade when visualizing the AOI	Millisecond
Peak velocity of saccades	The highest velocity reached during the saccade when visualizing the AOI	Degree per second
Pupil entropy	Predictability of pupil change	Bit

is more akin to a solo performance rather than teamwork, which is required during an actual surgery, eye behavior during communication and cooperation between team members was omitted in this study. Therefore, this study is unable to fully reveal the surgeons' eye behaviors during the entire hip arthroscopic surgery. Future studies should measure the complete set of surgeons' eye metrics for hip arthroscopy using different simulators for assessing operational performance and inter-team communication.

4. Conclusion

In this simulation-based training, medical trainees with different experience levels showed differences in eye behavior when performing hip arthroscopy. The results of this preliminary study demonstrate a good correlation between eye tracking metrics and participants' prior clinical experience. A higher level of surgical proficiency resulted in a lower cognitive load. Our findings demonstrated the possibility of applying eye tracking metrics data as part of an assessment tool to evaluate medical trainees' technical skills and establishing a benchmark for enhancing simulation-based training programs. Future work should focus on investigating the varied eye tracking metrics with larger sample sizes in simulated and real operational settings; thus, providing more valuable information on the pattern of eye movements during arthroscopic procedures.

5. Material and methods

This study was conducted at a single academic training center and was reviewed and approved by the university's institutional review board.

5.1. Participants

Participants of this study were orthopedic surgeons with prior experience in performing arthroscopic surgery independently and their surgical experience levels varied. They enrolled in this study voluntarily and received no remuneration or other benefits for their participation. Information regarding their training and prior clinical experience in arthroscopy was collected. Based on the participants' arthroscopic surgical experience, they were divided into three subgroups: novice, intermediate, and experienced groups. The surgical volume for the novice group was typically lower than other participants. For the intermediate and experience groups, the main distinguishing factor among these two groups was prior experience in performing hip arthroscopy, which would greatly influence the surgeons' behavior during the simulated operation. Before conducting the study, participants were informed of the purpose and design of this study, and could withdraw at any time without consequence. Consent was then obtained. Data collection was completed in July 2021.

5.2. Tasks and measures

In this study, a task-specific trainer was used for training in hip arthroscopic procedures (Figure 2A and B). Important surgical skills for hip arthroscopy, such as portal placement, intra-joint anatomical identification, and navigation could be performed using this simulator [39, 40, 41, 42]. Most equipment and tools used in this simulation-based training were identical to those of an actual operation. An arthroscopy demonstration linked to a laptop was used as the video system for the simulated procedure. Before the simulated operation, a 4-h didactic lecture on hip arthroscopy and operational instructions relating to the training content were given to all participants. During the simulated operation, participants were required to establish the anterolateral portal, mid-anterior portal, proximal mid-aterritorial portal, and distal anterolateral portal for hip arthroscopy. They were free to use fluoroscopy to assist with portal placement. Additionally, participants were required to identify the 12 o'clock, 2 o'clock, and 4 o'clock positions on the labrum.

The eye metric data of the simulated operation were recorded using the Tobii Pro Glasses 2 (Tobii Technology AB, Danderyd, Sweden), a portable eye tracking device (set at 50 Hz). This device has a head unit that embeds a video camera, microphone, and six infrared sensors. The video camera located at the center of the external side of the glass frame recorded the simulated operation. Infrared sensors internally mounted on the glass frame recorded eye movements and pupil behavior. The microphone recorded the audio information from the training session. Recordings were annotated using the Tobii Pro Lab Software (Tobii Technology AB). AOIs for this study were defined as the operative site, and arthroscopic and fluoroscopic fields (Figure 2C). The operative site referred to the simulator that was designed to replicate the soft tissue and bony structures of the hip joint. The arthroscopic field denoted the screen in the video system which simulated the visual images obtained from an arthroscopic camera. The captured images provided essential information for the intra-joint operation. The fluoroscopy screen displayed X-ray images that assisted portal placement. Participants' gaze direction and eye metrics were measured based on these AOIs. The gaze fixations were initially identified using the AOI selection tool of the Tobii Pro Lab Software (Tobii Technology AB) and manually verified by the research team.

Based on the specific setting of this study and available data sets, the data evaluated in this study included dwell time, number of fixations, number of saccades, saccade duration, peak velocity of saccade, and pupil entropy. The description of the metrics and unit used are shown in Table 9. Furthermore, the number of saccades reflected the switching times for fixation. Pupil entropy indicated the predictability of pupil change, or change in pupil size.

After the simulated operation, a modified NASA Task Load Index (NASA-TLX) survey was used to assess participants' perceived workload [17]. The NASA-TLX was used to semi-quantitatively measure the opinions of participants based on six subscales, including mental demands, physical demands, temporal demands, performance, effort, and frustration [43]. In this present study, each subscale comprised a continuous scale ranging from 0 (very low) to 10 (very high). These sub-scores were added to obtain the final NASA-TLX workload score, and presented as a final value ranging from 0 to 60 to indicate the perceived workload of each participant [17, 44].

5.3. Statistical analysis

The Jonckheere-Terpstra test, a rank-based nonparametric test, was used to analyze eye movement data for the operative site and arthroscopic field, and NASA-TLX data of the three different groups of participants. Due to the limited number of participants in the experienced group ($n = 1$) who used fluoroscopy, eye movement data on the fluoroscopic field for the experienced group were excluded from analysis, and the Mann-Whitney U test was used to compare data from the novice and intermediate groups. All statistical analyses were performed using the SPSS Statistics software (version 23.0; IBM, New York, USA). Results were considered statistically significant when the P value was less than 0.05.

Declarations

Author contribution statement

Bohong Cai, PhD; Cheng Chen: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Na Xu, PhD: Conceived and designed the experiments; Analyzed and interpreted the data.

Shengfeng Duan: Performed the experiments; Analyzed and interpreted the data.

Jiahui Yi: Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data.

Boon Huat Bay, PhD; Fangyuan Shen: Analyzed and interpreted the data; Wrote the paper.

Ning Hu: Performed the experiments.

Peng Zhang: Analyzed and interpreted the data.

Jie Chen: Analysed and interpreted the data; Wrote the paper.

Funding statement

Bohong Cai was supported by Humanity and Social Science Youth Foundation of the Ministry of Education of China [21YJCZH002], the Science and Technology Research Program of Chongqing Municipal Education Commission [KJQN202101001], Chongqing Higher Education Innovative Research Group [CXQT21026].

Shengfeng Duan was supported by "Chongqing and Chengdu Twin City Economic Circle Development" Technology and Innovation Program [KJCXZD2020034].

Ning Hu was supported by Chongqing Technology Innovation and Application Program [cstc2020jscx-msxmX0094].

Data availability statement

Data included in article/supp. material/referenced in article.

Declaration of interest's statement

The authors declare no competing interests.

Additional information

No additional information is available for this paper.

Acknowledgements

The authors wish to thank all participants for their involvement and are grateful to Chee San Tong for the language assistance. The authors appreciate Tobii China for providing the eye tracking devices that were used in this study.

References

[1] F.Y. Zhao, H.M. Dong, H.J. Huang, et al., The application of arthroscopic techniques in the diagnosis and treatment of hip-related conditions in China, Orthop. Surg. 13 (6) (2021) 1697–1706.
[2] G.J.M. Tuijthof, M.N. van Sterkenburg, I.N. Sierevelt, J. van Oldenrijk, C.N. Van Dijk, G.M.M.J. Kerkhoffs, First validation of the PASSPORT training environment for arthroscopic skills. Knee Surgery, Sport Traumatol. Arthrosc. 18 (2) (2010) 218–224.
[3] A. Alvand, S. Auplish, H. Gill, J. Rees, Innate arthroscopic skills in medical students and variation in learning curves, J. Bone Jt. Surg.-Am. 93 (19) (2011) e115(1)–e115(9).
[4] A.J. Carr, A.J. Price, S. Glyn-Jones, J.L. Rees, Advances in arthroscopy - indications and therapeutic applications, Nat. Rev. Rheumatol. 11 (2) (2015) 77–85.
[5] N. Mehta, P. Chamberlin, R.G. Marx, et al., Defining the learning curve for hip arthroscopy: a threshold analysis of the volume-outcomes relationship, Am. J. Sports Med. 46 (6) (2018) 1284–1293.
[6] G.D. Dumont, R.M. Cohn, M.M. Gross, T.J. Menge, N.C. Battle, Z.T. Thier, The learning curve in hip arthroscopy: effect on surgical times in a single-surgeon cohort, Arthrosc. J. Arthrosc. Relat. Surg. 36 (5) (2020) 1293–1298.
[7] R. Aggarwal, Just-in-time simulation-based training, BMJ Qual. Saf. 26 (11) (2017) 866–868.
[8] P. Bradley, The history of simulation in medical education and possible future directions, Med. Educ. 40 (3) (2006) 254–262.
[9] L. Phillips, J.J.H. Cheung, D.B. Whelan, et al., Validation of a dry model for assessing the performance of arthroscopic hip labral repair, Am. J. Sports Med. 45 (9) (2017) 2125–2130.
[10] T.C.B. Pollard, T. Khan, A.J. Price, H.S. Gill, S. Glyn-Jones, J.L. Rees, Simulated hip arthroscopy skills: learning curves with the lateral and supine patient positions, J. Bone Jt. Surg. 94 (10) (2012) e68.
[11] D.E. Bauer, K. Wieser, A. Aichmair, P.O. Zingg, C. Dora, S. Rahm, Validation of a virtual reality–based hip arthroscopy simulator, Arthrosc. J. Arthrosc. Relat. Surg. 35 (3) (2019) 789–795.
[12] M. Morgan, A. Aydin, A. Salih, S. Robati, K. Ahmed, Current status of simulation-based training tools in orthopedic surgery: a systematic review, J. Surg. Educ. 74 (4) (2017) 698–716.
[13] A. Cuschieri, N. Francis, J. Crosby, G.B. Hanna, What do master surgeons think of surgical competence and revalidation? Am. J. Surg. 182 (2) (2001) 110–116.
[14] R. Aggarwal, T.P. Grantcharov, A. Darzi, Framework for systematic training and assessment of technical skills, J. Am. Coll. Surg. 204 (4) (2007) 697–705.
[15] K. Moorthy, Y. Munz, S.K. Sarker, A. Darzi, Objective assessment of technical skills in surgery, Br. Med. J. 327 (7422) (2003) 1032–1037.
[16] R.J. Koehler, S. Amsdell, E.A. Arendt, et al., The arthroscopic surgical skill evaluation tool (ASSET), Am. J. Sports Med. 41 (6) (2013) 1229–1237.
[17] C. Wu, J. Cha, J. Sulek, et al., Eye-tracking metrics predict perceived workload in robotic surgical skills training, Hum. Factors 62 (8) (2020) 1365–1386.
[18] L. Richstone, M.J. Schwartz, C. Seideman, J. Cadeddu, S. Marshall, L.R. Kavoussi, Eye metrics as an objective assessment of surgical skill, Ann. Surg. 252 (1) (2010) 177–182.
[19] C. Pawelke, F. Merkle, D. Kurtovic, S. Gierig, G. Müller-Plath, Comparison of a Perfusion Simulator to a Clinical Operating Room: Evaluation of Eye Tracking Data and Subjective Perception. A Pilot Study, Perfus (United Kingdom). Published online, 2020.
[20] T. de Greef, H. Lafeber, H. van Oostendorp, J. Lindenberg, Eye movement as indicators of mental workload to trigger adaptive automation, in: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Vol. 5638 LNAI, 2009, pp. 219–228.
[21] G. Marquart, C. Cabrall, J. de Winter, Review of eye-related measures of drivers' mental workload, Procedia Manuf. 3 (Ahfe) (2015) 2854–2861.
[22] K.A. Ericsson, An expert-performance perspective of research on medical expertise: the study of clinical performance, Med. Educ. 41 (12) (2007) 1124–1130.
[23] T. Tien, P.H. Pucher, M.H. Sodergren, K. Sriskandarajah, G.Z. Yang, A. Darzi, Eye tracking for skills assessment and training: a systematic review, J. Surg. Res. 191 (1) (2014) 169–178.
[24] M.M. Shahimin, A. Razali, An eye tracking analysis on diagnostic performance of digital fundus photography images between ophthalmologists and optometrists, Int. J. Environ. Res. Publ. Health 17 (1) (2020).
[25] N. Herten, T. Otto, O.T. Wolf, The role of eye fixation in memory enhancement under stress – an eye tracking study, Neurobiol. Learn. Mem. 140 (2017) 134–144.
[26] C.M. Schulz, E. Schneider, L. Fritz, et al., Eye tracking for assessment of workload: a pilot study in an anaesthesia simulator environment, Br. J. Anaesth. 106 (1) (2011) 44–50.
[27] X. He, L. Wang, X. Gao, Y. Chen, The eye activity measurement of mental workload based on basic flight task, in: IEEE International Conference on Industrial Informatics (INDIN), 2012, pp. 502–507.
[28] I. Škvareková, P. Pecho, V. Ažaltovic, B. Kandera, Number of saccades and fixation duration as indicators of pilot, Transport. Res. Procedia 51 (2019) (2020) 67–74.
[29] P.J.A. Unema, S. Pannasch, M. Joos, B.M. Velichkovsky, Time course of information processing during scene perception: the relationship between saccade amplitude and fixation duration, Vis. Cogn. 12 (3) (2005) 473–494.
[30] S. Pannasch, B.M. Velichkovsky, Distractor effect and saccade amplitudes: further evidence on different modes of processing in free exploration of visual images, Vis. Cogn. 17 (6–7) (2009) 1109–1131.
[31] B.M. Velichkovsky, A. Rothert, M. Kopf, S.M. Dornhöfer, M. Joos, Towards an express-diagnostics for level of processing and hazard perception, Transport. Res. F Traffic Psychol. Behav. 5 (2) (2002) 145–156.
[32] M. Mills, A. Hollingworth, S. Van der Stigchel, L. Hoffman, M.D. Dodd, Examining the influence of task set on eye movements and fixations, J. Vis. 11 (8) (2011) 17.
[33] L.L. Di Stasi, A. Catena, J.J. Cañas, S.L. Macknik, S. Martinez-Conde, Saccadic velocity as an arousal index in naturalistic tasks, Neurosci. Biobehav. Rev. 37 (5) (2013) 968–975.
[34] G.G. Menekse Dalveren, N.E. Cagiltay, Insights from surgeons' eye-movement data in a virtual simulation surgical training environment: effect of experience level and hand conditions, Behav. Inf. Technol. 37 (5) (2018) 517–537.
[35] V. Bachurina, M. Arsalidou, Multiple levels of mental attentional demand modulate peak saccade velocity and blink rate, Heliyon 8 (1) (2022).
[36] T. Tien, P.H. Pucher, M.H. Sodergren, K. Sriskandarajah, G.Z. Yang, A. Darzi, Differences in gaze behaviour of expert and junior surgeons performing open inguinal hernia repair, Surg. Endosc. 29 (2) (2015) 405–413.
[37] M.S. Martin, B. Huard-Nicholls, A.P. Johnson, Gaze and pupil size variability predict difficulty-level and safe intersection crosses in a driving simulator, Proc. Hum. Factors Ergon. Soc. Annu. Meet. 65 (1) (2021) 843–847.
[38] B. Zheng, M.A. Cassera, D.V. Martinec, G.O. Spaun, L.L. Swanström, Measuring mental workload during the performance of advanced laparoscopic tasks, Surg. Endosc. 24 (1) (2010) 45–50.
[39] J. Kautzner, P. Zeman, A. Stančák, V. Havlas, Hip arthroscopy learning curve: a prospective single-surgeon study, Int. Orthop. 42 (4) (2018) 777–782.
[40] M. Karahan, P. Randelli, G.M.M.J. Kerkhoffs, G.J.M. Tuijthof, E. Asbl, Effective training of arthroscopic skills, in: Effective Training of Arthroscopic Skills, 2015, pp. 1–173.
[41] Y. Hui, O. Safir, A. Dubrowski, H. Carnahan, What skills should simulation training in arthroscopy teach residents? A focus on resident input, Int. J. Comput. Assist. Radiol. Surg. 8 (6) (2013) 945–953.
[42] L. Haipeng, L. Ji, Z. Juanli, et al., Portal setup: the key point in the learning curve for hip arthroscopy technique, Orthop. Surg. 13 (6) (2021) 1781–1786.
[43] S.G. Hart, L.E. Staveland, Development of NASA-TLX (task load index): results of empirical and theoretical research, Adv. Psychol. 52 (C) (1988) 139–183.
[44] S.G. Hart, Nasa-task load index (NASA-TLX); 20 years later, Proc. Hum. Factors Ergon. Soc. Annu. Meet. 50 (9) (2006) 904–908.