Introduction

The expert-novice apprenticeship model is a critical part of vocational learning for a variety of visual-intensive professions, such as surgery, air traffic control, and forensic analysis (Köpsén & Nyström, 2012; Koskela & Palukka, 2011; Walter, 2006). Upon completion of a relevant degree, a novice may enter an expert-novice apprenticeship lasting several years; for example, medical residency programmes and forensic apprenticeships can take (up to) five years to complete. Current research into visual expertise development typically involves one of two approaches. Firstly, comparing the processing behaviours and performance of experts with those of novices (Gegenfurtner et al., 2011) such as comparing expert and novice forensic face examiners or strategies for approaching crime scenes (Baber & Butler, 2012; White et al., 2015). Secondly, creating instructional materials for novices based on expert behaviours, including the development of eye movement modelling examples (EMMEs), an instructional video technique displaying a replay of experts' eye movements facilitated by the advancements in eye-tracking technology (Emhardt et al., 2023; van Gog et al., 2008; Valtakari et al., 2021). With the existing focus on expert-novice comparisons and learning through instructional video, a domain that is unexplored and further deserves attention is direct didactic interactions, in other words, scaffolded interactions (Jarodzka et al., 2020; Nückles, 2021).

Experts exhibit scaffolding behaviours during direct interactions with novices, resulting in extraneous information being offloaded from the novice's working memory to improve comprehension and learning (Van Nooijen et al., 2024). Such scaffolding behaviours can be classified into two overarching categories based on their function relative to the working memory. *Cueing* behaviours by experts are defined as scaffolding behaviours (verbal or nonverbal) that can reduce the incoming information entering a novice's working memory, reducing the informational load of added information on the novice's cognitive capacity. Examples include experts pointing at or prompting novices with relevant information, for example, "click the menu button in the top left corner". *Chunking* behaviours by experts are defined as scaffolding behaviours that reduce cognitive load for novices by grouping multiple pieces of new or existing information, freeing up space in the working memory. Grouping information together, for example into cognitive schemas, can improve retention and retrieval (Gobet et al., 2001). Examples include highlighting sequences/order of events, such as explaining that one tool is always used first before use of a second tool, or use of metaphors, illustrated by Sutkin and colleagues' observation of surgeons: *"That's what I tell everyone in the operating room: the pelvis is a bowl and you're shaking hands to open up your spaces"* (2015, p. 249).

Scaffolding behaviours like cueing and chunking are often defined from the perspective of an expert scaffolding a novice's learning. However, we argue that novices can also use cueing and chunking behaviours in dyadic interactions. The function of such 'scaffolding' behaviours is twofold. The first is as a metacognitive tool that helps make their needs explicit, e.g. novice chunking could be asking clarifying questions like "why" or seeking ways to group information in their working memory to reduce their cognitive load. The second is as a reverse 'scaffold' for the expert's teaching process, providing experts with information to help them be more contingent in their scaffolding, e.g. novice cueing could be nodding their head or saying "I understand this part" to signal to experts they can move on, reducing the amount of extraneous information experts introduce to the novice's working memory. The present research will therefore qualitatively code both cueing and chunking exhibited by the expert, as well as cueing and chunking exhibited by the novice (Van Nooijen et al., 2024).

Emerging evidence suggests that experts modify both their scaffolding behaviours and gaze while guiding novices through visual problem-solving tasks with visualisations (Emhardt et al., 2020; Jaarsma et al., 2018). Research by Jaarsma and colleagues (2018) showed that verbal contributions by novices increase when the novice has control over the visual task. The type of verbal content also changes depending on who is in control. Shvarts and Abrahamson (2019) provide insights into tutor-student joint visual attention (JVA), as JVA can be analysed moment-to-moment within qualitative small-sample research. JVA, which occurs when the gaze of two or more individuals converges on the same point within a shared visual environment, is a predictor of learning and performance (Becker et al., 2021; Schneider et al., 2015; Schneider & Bryant, 2024). Shvarts and Abrahamson (2019) highlight the importance of coupling of visual attention in educational settings as a mechanism for achieving mutual understanding and coordinating actions towards a common goal. They indicate that this coupling is facilitated through both verbal and non-verbal scaffolding behaviours, including gestures, expressions, and the use of shared materials or tools, which serve as mediators of cognitive processes. JVA may influence performance via interaction quality (Sharma et al., 2020); JVA may be an indicator of successful interaction quality (Jermann et al., 2011). In dyads who can see their partners' gaze in real-time, increased JVA also leads to improved collaboration quality (Schneider & Pea, 2013).

Besides JVA, reciprocal gaze has also been shown to improve student learning in various educational settings (Dalzel-Job et al., 2011; Fiorella et al., 2019; Schneider & Pea, 2016). Eye contact from the novice may signal engagement, whereas eye contact from the expert may serve to gauge the novice's engagement, or cues such as facial expressions that indicate confusion (Kendon, 1967). Given the important and time-sensitive role of contingency in scaffolding, facial feedback obtained during reciprocal eye contact may help experts make their instruction more contingent on the novice's needs, potentially leading to better-timed scaffolding and subsequently improved learning and performance for the novice. Experts may also initiate reciprocal eye contact as the result of their intent to scaffold a student. Haataja and colleagues (2019) demonstrate in a case study that a teacher directs their gaze to their students' faces most during the application of affective scaffolding (improving student motivation and reducing frustration), whereas during cognitive scaffolding (bridging student competence with the task) they direct their gaze to the learning material. Reciprocal gaze has been studied in various educational settings with trained teachers; understanding how and when reciprocal gaze occurs with experts without pedagogical training (and how it relates to JVA) can add a new layer of insight to our understanding of expert-novice learning is scaffolded, face-to-face interactions.

However, capturing these interactions is easier said than done. Although perceptual measures like eye tracking are growing in popularity, such technologies can be difficult to use in combination with traditional expert-novice scenarios; for example, screen-based eye-trackers can only capture gaze from one person at a time. There is a growing need for measurement and analysis methods that allow traditional interactions like scaffolding behaviours to be studied in combination with gaze metrics like joint visual attention (JVA), in naturalistic expert-novice scenarios. Both reciprocal gaze and JVA can be measured with dual wearable eye tracking. Measuring JVA in collaborative settings has been done using wearable eye tracking in combination with fiducial markers, as detailed in Hessels et al. (2023) and Schneider (2020). Using fiducial markers allows for the gaze points of two or more individuals using wearable eye-trackers to be mapped onto the same area, providing insight into how close together or far apart the gaze points are. This approach provides a structured way to contrast scaffolding behaviour during moments of high and low joint visual attention. We aim to connect the presence of high or low JVA with scaffolding behaviours exhibited during the expert-novice interaction, such as cueing or chunking. To our knowledge, this research is the first to bridge JVA with concrete scaffolding actions in expert-novice interaction research.

The aim of the present research is thus twofold; to identify the use of cueing and chunking in expert-novice interactions, and to provide a preliminary exploration into the connection between these cueing and chunking behaviours and JVA using fiducial markers and a contrast-based analysis method. This is done in a learning scenario in the experts' workplace, where the expert first demonstrates to the novice how to complete the task (expert control), followed by the novice attempting the task themselves (novice control). Although no statistical analyses will be applied in the present research due to the small sample size, hypotheses are stated for qualitative findings and descriptive statistics.

1. More verbal and gestural cueing and less reciprocal gaze are anticipated in segments with high JVA than segments with low JVA, as cueing can lead to JVA.
2. More verbal and gestural chunking and more reciprocal gaze are anticipated in segments with low JVA than segments with high JVA, as chunking is conceptually more abstract and does not necessitate JVA.

Given the prevalence and high-stakes nature of forensic professions where face comparison and feature comparison expertise are required, years of repeated exposure may not be feasible or optimal. Understanding the scaffolding behaviours that forensic experts use to communicate their knowledge, and the relationship these behaviours have with the expert-novice JVA may lead to insights that can help improve both our theoretical understanding of apprenticeship-style expert-novice interactions in visual tasks and potentially provide a basis for future research into how expert-novice interactions can be optimised, leading to improved communication of visual expertise in forensic domains and beyond.

Methods

Participants and design

Expert participants were three registered experts currently working at the Netherlands Forensic Institute, having worked within their current team for an average of 18.33 years ($\sigma = 5.25$). The expert team we recruited from has eight members of which roughly six work full time in this team.

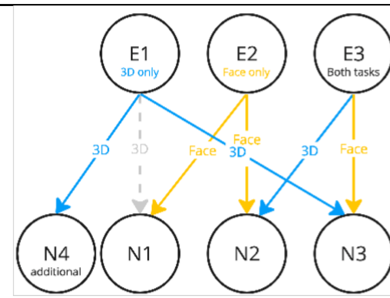
Novice participants consisted of four female bachelor Psychology students in the third year of their studies with an average age of 21.25 years ($\sigma = 1.26$). Novice participants had no prior experience with forensic tasks. Novice participants were awarded study credits as compensation for their participation.

This research employed a comparative multiple case study design with expert-novice pairings described in Table 1. Initially three novice participants were recruited, however due to a recording error during the first novice recording, this data could not be analysed and an additional fourth novice participant was recruited.

Table 1

Participant Pairing and Learning Task Per Session

Case	Expert	Novice	Task
*0***	*Expert 1 ***	*Novice 1 ***	*Feature comparison ***
1	Expert 1	Novice 3	Feature comparison
2	Expert 1	Novice 4	Feature comparison
3	Expert 2	Novice 1	Face comparison
4	Expert 2	Novice 2	Face comparison
5	Expert 3	Novice 2	Feature comparison
6	Expert 3	Novice 3	Face comparison



Note. Expert 2 had more expertise in the face comparison task than the feature comparison task, and vice versa for Expert 1, so it was chosen to have these experts teach the task they had the most expertise in. Expert 3 worked with both tasks. **Due to an incomplete recording, this session was excluded and re-recorded with the same expert and a new novice participant Novice 4.

Materials

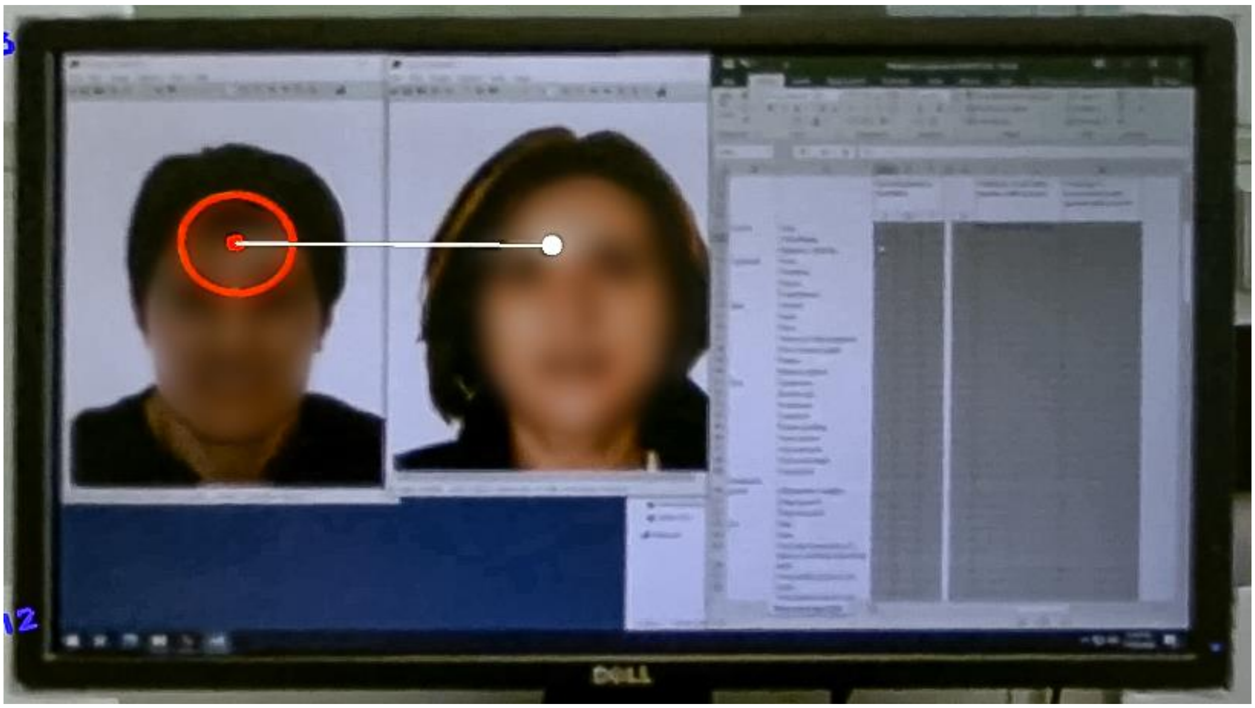
Two visual tasks were selected for the present research that both accurately reflect the expertise of our expert participants, and that we anticipate will be reflective of domain-general comparative visual expertise. These tasks are face comparison and feature comparison. Face comparison is an umbrella term for tasks comparing (unfamiliar) faces in real life and images; face comparison experts can identify unfamiliar faces more accurately than novices (Kroon, 2017; Phillips et al., 2018). In practice, this translates to high-stakes professional tasks including passport identification during border control or CCTV analysis by security experts. Feature comparison research encompasses fields like handwriting analysis, security footage analysis or fingerprint analysis. Repeated exposure to structured analysis of unfamiliar faces or features over time, whether through training or on-the-job experience, is currently the most effective existing technique to gain expertise (Beebe & Guynes, 2006; Sporer, 2022; Wilkinson & Evans, 2008). Recent research provides evidence that different domain-specific comparative tasks conducted by visual experts elicit similar psychological processes (Growns et al., 2022). We thus anticipate that our results using these tasks will be generalisable to comparative visual expertise in non-forensic visual domains.

For the feature comparison, the expert participants selected cases processed by their team in the past that were both completed and permitted for research use. For the face comparison task, materials from a collaborative exercise used for professional alignment between experts were used.

Face comparison

Figure 1

Face Comparison Task Stimulus



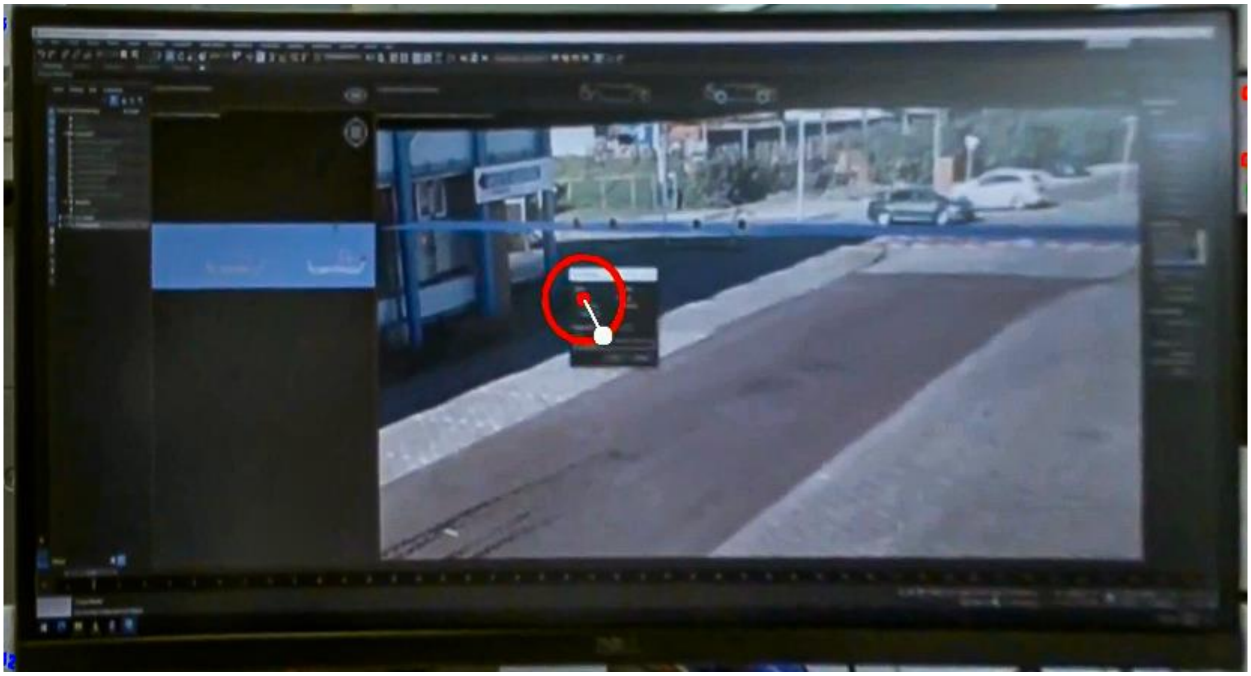
Note. Screenshot from data analysis showing the two faces for comparison on the left, and the structural checklist on the right. All windows are adjustable, allowing participants to expand/contract window size or zoom in/out on facial features. The red dot shows the gaze location of the expert, the white dot shows the gaze location of the novice.

The face comparison task consisted of two photographs of a female face, and a checklist spreadsheet used by the experts in practice. The checklist highlights several facial features and characteristics of the image quality, which the novice must structurally compare between the two photographs, culminating in a specific evaluation (same person or not the same person) at the end. To evaluate the novices' ability, the conclusion drawn by novices are compared to that of an expert.

Feature comparison

Figure 2

Feature Comparison Task Stimulus



Note. Screenshot from data analysis showing the 3DS Max software. A second screen was set up next to this screen with a still from the original security footage, used as reference during the camera match portion of the task. The red dot shows the gaze location of the expert, the white dot shows the gaze location of the novice.

The feature comparison task consisted of a file with a 3D car set up in the software Autodesk 3ds Max (Harper, 2012), along with stills from a surveillance camera to be used as reference. First a ‘camera match’ must be made, where points (features) in the 3D environment must be matched with points from the surveillance footage, and subsequently a virtual car must be placed and moved in the environment. The car is placed a total of 10 times using the ‘matched’ camera perspective. To evaluate the novices’ ability, the deviation of the novices’ car placement in the 3D software relative to the true position of the car in the reference stills are measured (this is the perceptual comparison). Their standard deviation value is compared to that of an expert.

Task phases

Both tasks were set up to contain two phases; an expert control case, where the expert explained the content to a listening novice through demonstration, and a novice control case, where the novice completed the case under the guidance of the expert. Although initially, the aim was to have a controlled post-test where novices complete the task independently, piloting the tasks revealed that both tasks were too difficult to master within the span of a single learning session. Therefore, we opted for expert-control and novice-control phases, like the conditions by Jaarsma and colleagues (2018), with the addition that the experts were instructed to try not to help the novices in the novice-control phase unless the novice was stuck.

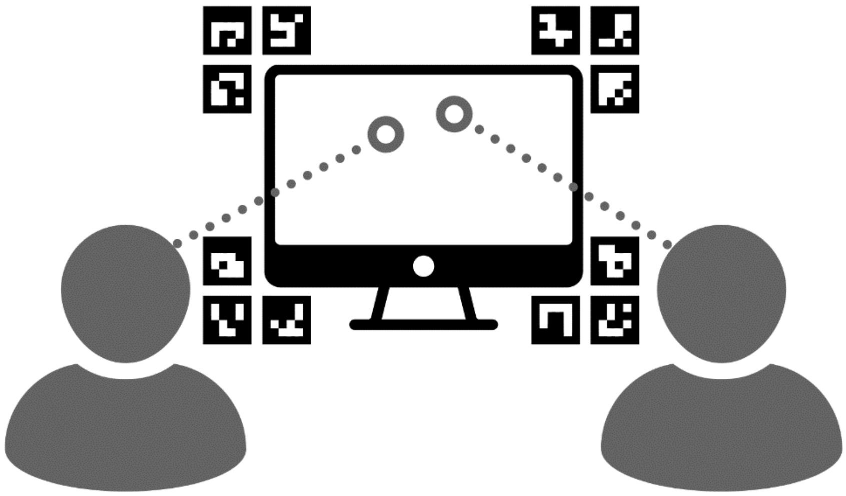
Apparatus

Both the expert and novice participants wore Tobii Pro 3 eye-tracking glasses for the duration of each task, recording gaze and video (including audio) using Glasses 3 Controller (Tobii, 2023a). The Tobii Pro 3 glasses have a scene camera resolution of 1920x1080, recording at 25fps, with a diagonal field of view of 106° (Tobii, 2023b). The participants were seated at the expert's desk during the tasks. A video recording was taken from behind the participants, which served as a back-up in case participant movements or speech were unclear from the eye-tracking recordings alone. The glassesValidator method was used to determine the accuracy of the eye-tracking recordings (see "Data quality" for more information; Niehorster et al., 2023).

ArUco markers were adhered to the corners of the experts' screen, as shown in Figure 1. After piloting, we chose to use 12 markers (three per corner) instead of four to provide more stability to the analysis in case the participants sat in a way such that the corners of the screen were only partially visible.

Figure 3

ArUco Marker Positioning Around the Screen for JVA Analysis



Note. ArUco markers 1-12 were used in the present research.

Procedure

Participants were given the informed consent forms to review and sign prior to their participation. Next, participants were fitted with Tobii eye-tracking glasses. The glasses were calibrated using the Tobii bullseye calibration method. Next, both participants were guided through the GlassesValidator poster (targets 1-9) by the researcher. The poster was placed at the same distance as the computer screen. The expert control phase of the research then began, where the expert explained how to complete the task. When the expert and novice both indicated they were ready to move on, the expert opened a new file, and the novice was then allowed to attempt the task (novice control phase). Experts guided the novices where necessary. After the end of this phase, the GlassesValidator poster was used a second time to account for eye-tracker slippage, before ending all recordings and closing the session.

Analyses

Data quality

We included gaze data with at least 80% recorded gaze samples during the learning phase. The glassesValidator poster was used to determine the accuracy of the gaze data as measured with the eye-tracking glasses at the beginning and end of each session (Niehorster et al., 2023). We included participants if the accuracy for both participants was lower than ±3° both at the start and the end of the recording. Note that no existing reporting guidelines specify minimum thresholds for gaze samples and accuracy; the present thresholds were selected after piloting. Recorded data quality for each case can be seen in Table 2. All accuracy values were within the specified range of ±3°.

Table 2

Gaze Accuracy Across Six Recorded Cases

	FA 1		**FA 2**		**FA 3**		**FE 1**		**FE 2**		**FE 3**	
	Exp	Nov	Exp	Nov	Exp	Nov	Exp	Nov	Exp	Nov	Exp	Nov
Accuracy start °	1.135	1.181	0.838	0.549	0.785	0.547	0.419	1.049	0.345	1.196	0.654	0.496
Accuracy end °	0.538	1.504	0.863	0.823	2.251	1.175	0.310	1.121	0.380	1.637	0.636	1.523
Length hh:mm:ss	00:38:05		00:34:55		00:39:22		00:37:28		00:45:57		01:14:28	

Note. Accuracy determined using the GlassesValidator poster; acc_2D over 9 targets. Length indicates the total learning time, which is the time between the end of the first GlassesValidator measurement, and the start of the second GlassesValidator measurement.

Data preprocessing

Eye-tracking video files. Video files were exported for each participant using Tobii Glasses 3 Controller. The integrated export saves the participants' scene camera footage with an overlayed red circle indicating the participants' gaze. No further filters or adjustments were applied beyond those that occur standard in the software. Although video recording had been manually started at the same time for each dyad, a few frames delay was still notable in the videos, so instead the videos were trimmed by hand to line up at two points on the same frame. All videos had the same frame rate upon export (manufacturer framerate 25fps, true framerate 24.92fps), however the videos in case 4 were misaligned by ~1 second near the end of the recording (expert video was slower) despite starting on the same frame. The expert video was manually sped up by 1 second to mitigate this difference, however some minor misalignment is still evident at some timepoints in the integrated video. This was manually accounted for by the researchers when coding reciprocal vs. delayed gaze in case 4. As described in the next section, JVA was averaged over a range of 1-10 seconds, therefore the misalignment of <1 second in case 4 was deemed acceptable for the analysis.

Joint visual attention. A script was written in Python using OpenCV (Bradski, 2000) to quantify joint visual attention in each case. The python script used the expert and novice Tobii glasses 3 controller export videos as input, analysed the gaze point, used the ArUco markers to projective-transform the image so they could be overlayed, then calculated the Euclidean distance between the expert and novice gaze (JVA). This analysis was done per frame and exported both as integrated video and a .csv file with JVA per frame.

Analyses

Establishing Segments with High and low JVA. A second script was written in Python using OpenCV (Bradski, 2000) to analyse the moments of highest distance between expert and novice gaze points and lowest distance between expert and novice gaze points. This allowed us to complete a contrast-based qualitative analysis, where we contrast coded elements in segments of high JVA with segments of low JVA. To do this, the JVA per-frame data was first smoothed by averaging data points per 1 second. Next, 20 timepoints of highest and lowest JVA were identified. If points were within < 4 s of each other, they were combined and the next point of highest/lowest JVA is included. This process was repeated for smoothing steps of 2, 3, through to 10 s. Finally, the high/low JVA sets over different smoothing steps were compared, and values that occur at a point of high/low JVA in at least 2 smoothing step sets are included (ranked on most frequent occurrence). The 20 most occurring values for high/low JVA across 10 smoothing steps are identified and used to cutting the video segments at 5 s before/after the value, leading to 20 high and 20 low JVA segments of 10 seconds each per video. These segments were used in qualitative coding. The final dataset included 240 segments, including the 20 highest and 20 lowest JVA segments per case. This means all six cases are equally represented in number of segments, allowing for descriptive statistics to be reported alongside qualitative insights. In the event a segment was not deemed fit for qualitative analysis, this segment was discarded from the dataset and the next segment of highest or lowest JVA was adopted instead. Ultimately, a total of 263 segments were analysed, of which 23 were removed due to technical reasons such as incomplete marker detection, or content reasons such as questions being asked to the researcher about the eye-trackers, leaving 240 for coding.

Qualitative coding. The coding scheme in Appendix 1 was used to code each segment on the following dimensions: presence of reciprocal gaze, expert gaze towards novice, novice gaze towards expert, expert cueing, expert chunking, expert verbal behaviour, expert nonverbal behaviour, novice cueing, novice chunking, novice verbal behaviour, novice nonverbal behaviour, and the presence of any findings (up to 3) from the list of 12 expert-novice interaction findings from van Nooijen and colleagues (2024). These coding dimensions were constructed for the present research, the cueing and chunking flowcharts from Van Nooijen et al. (2024), and feedback from two independent raters who used a preliminary version of this scheme during an initial pilot. All segments were coded by the primary researcher, with 10% (24) segments coded by another researcher to determine the inter-rater reliability for each code. A minimum Cohen's kappa value of 0.6 was required (indicating moderate agreement as presented in McHugh, 2012); in the event this value was not reached, the researchers discussed definition refinements and re-coded the segments. The minimum threshold of 0.6 was reached for all codes as shown in Appendix 2.

Transparency and Openness

We describe our participant recruitment, data exclusions, all materials, and all measures in the study, and we adhered to the Journal of Applied Psychology methodological checklist. Python scripts are created using the open access Python library OpenCV and are available upon request from the first author. Research materials are described in full and are not available due to their proprietary nature. Both raw and processed data for this research contain biometric and identifying personal information, meaning they are not freely available for reuse; data is stored in a secure online repository for 10 years after publication. Ethics approval was gained for this research. No preregistration was completed for this research.

Results and Discussion

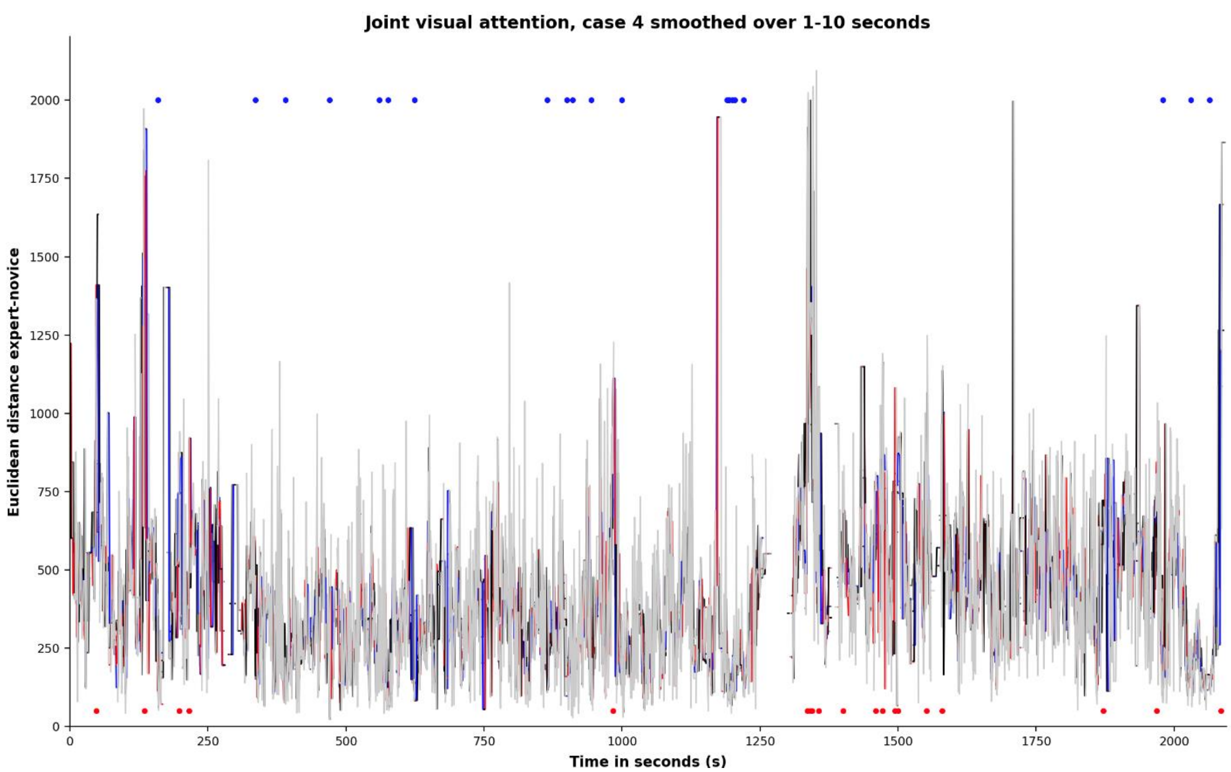
In the following section, the descriptive and qualitative outcomes of the present study will be presented and discussed in the context of related research into expert-novice interactions.

Total JVA across each case

Below, Figure 4 illustrates in a randomly selected case how joint visual attention fluctuates over time, smoothed over 1-10 seconds. Points of highest (blue) and lowest (red) JVA are plotted for reference; these points comprise the segments used in our qualitative analysis. An absence of data means the participants were not both looking at the screen and were instead for example looking at each other or at the keyboard/mouse. Overall trends in JVA are often representative of changes in the task or overall approach; moments where the JVA becomes higher and more consistent often reflect the expert monitoring novice behaviour, whereas more missing data in the graph may indicate a higher amount of reciprocal eye contact. No differences in trends are seen between the two task types.

Figure 4

JVA in Case 4



Descriptive and qualitative segment analysis

Cueing and chunking scaffolding behaviours

Table 3

Cueing and Chunking Scaffolding Behaviours Across Six Recorded Cases

	Both tasks	FA 1	FA 2	FA 3	FE 1	FE 2	FE 3	Total	Face task	Feature task
Total expert cueing	**63.3%**	**29**	**16**	**31**	**25**	**26**	**25**	**152**	**50.0%**	**50.0%**
High JVA + verbal cueing	39.2%	15	8	19	14	19	19	94	55.3%	44.7%
Low JVA + verbal cueing	24.2%	14	8	12	11	7	6	58	41.4%	58.6%
High JVA + nonverbal cueing	31.3%	12	3	15	9	17	19	75	60.0%	40.0%
Low JVA + nonverbal cueing	14.6%	12	4	10	3	1	5	35	25.7%	74.3%
Total expert chunking	**39.2%**	**14**	**6**	**17**	**12**	**21**	**24**	**94**	**60.6%**	**39.4%**
High JVA + verbal chunking	17.9%	4	2	8	4	10	15	43	67.4%	32.6%
Low JVA + verbal chunking	21.3%	10	4	9	8	11	9	51	54.9%	45.1%
High JVA + nonverbal chunking	16.7%	2	1	8	4	10	15	40	72.5%	27.5%
Low JVA + nonverbal chunking	14.6%	10	3	8	6	2	6	35	40.0%	60.0%
Total novice cueing	**58.3%**	**25**	**36**	**22**	**23**	**17**	**17**	**140**	**40.7%**	**59.3%**
High JVA + verbal cueing	15.4%	7	6	6	10	6	2	37	48.6%	51.4%
Low JVA + verbal cueing	20.4%	6	5	10	12	10	6	49	57.1%	42.9%
High JVA + nonverbal cueing	18.8%	11	18	5	7	3	1	45	24.4%	75.6%
Low JVA + nonverbal cueing	24.6%	9	16	9	9	7	9	59	42.4%	57.6%
Total novice chunking	**3.3%**	**1**	**1**	**0**	**6**	**0**	**0**	**8**	**75.0%**	**25.0%**
High JVA + verbal chunking	1.7%	0	1	0	3	0	0	4	75.0%	25.0%
Low JVA + verbal chunking	1.7%	1	0	0	3	0	0	4	75.0%	25.0%
High JVA + nonverbal chunking	1.7%	0	1	0	3	0	0	4	75.0%	25.0%
Low JVA + nonverbal chunking	1.3%	1	0	0	2	0	0	3	66.7%	33.3%

Note. "Both task" percentages represent proportion of the 240 total segments. Face and feature comparison task percentages represent the proportion of "Both task" percentages.

As indicated in Table 3, (verbal) expert cueing was the most commonly occurring scaffolding behaviour, with 63.3% of all segments containing (verbal) expert cues, distributed equally between the two tasks. Expert cues occurred more often in segments of high JVA (39.2% verbal and 31.3% nonverbal) relative to segments of low JVA (24.2% verbal and 14.6% nonverbal), in line with Hypothesis 1. Expert chunking, on the other hand, was demonstrated by experts in 39.2% of segments, distributed relatively equally over task type and verbal/nonverbal expression. Many expert cues were mouse movements and gestures (often combined with verbal cues such as "look") that served to guide the attention of novices, which is likely to have affected JVA. During qualitative analysis we found many instances of experts using verbal (explanations) and nonverbal (demonstrations) cueing and chunking to clarify tool use, which occurred more frequently at points of high JVA. This is an expected finding as the tools in both tasks were on-screen, and in line with findings from Van Nooijen and colleagues (2024). Cues were also often used to provide concrete information such as landmarks, which often prompted actions from the novice. These cues were noted during segments of high JVA and noted more frequently during the feature comparison task than during the face comparison task. A possible explanation for this is that the face comparison task inherently contains cues. The task material, a checklist, is on screen throughout the task, allowing novices to follow the list step-by-step without needing expert assistance. The feature comparison task had no such on-screen 'guide'; novices thus probably required more expert guidance in that task.

It was hypothesized that during moments of low JVA, more chunking would occur than during moments of high JVA (Hypothesis 2), however no such distinction was found. It is possible that our hypothesised distinction between high and low JVA relative to the screen was too fine-grained, and that rather chunking was more likely to occur when participants are looking at each other, during moments of no JVA; future research could take a similar approach to reciprocal gaze, first coding all moments where reciprocal gaze occurs (e.g. using automated facial detection tools) and then subsequently investigate the type of scaffolding behaviours that take place during those moments.

Novice cueing was coded in an unexpectedly high percentage of segments (58.3%). In previous research, cueing was found to be one-way from the expert to the novice (Van Nooijen et al., 2024). However, novice cues often did not relate to the task content specifically but were short verbal or nonverbal prompts to the expert to continue; behaviours like nodding 'yes' to the expert as they explained, or short confirmatory phrases like "yes" or "okay". We consider those to be cues, as they function to avoid superfluous explanation from experts, and thus serve to avoid extraneous cognitive load

(load related to the way the task is presented). At the same time, these cues do not function to lower cognitive load by directing the attention of the other the way cues such as pointing do, and thus these cues do not directly affect JVA. This is in line with the presence of many “acknowledgements” from novices coded by Jaarsma and colleagues (2018). Contrary to findings from Van Nooijen and colleagues (2024), we did not just find that experts reassure novices to continue, but also the opposite: novices often reassure experts by prompting them to continue. Novice cueing also occurred more often in the feature comparison task than the face comparison task, especially nonverbal novice cues during moments of high JVA in the feature comparison task (75.6%) relative to the face comparison task (24.4%). Further research could investigate whether the presence or absence of such acknowledging cues directly impacts experts’ ability to provide contingent support on novice learners. Novice chunking occurred very rarely in the 240 segments (3.3%) often in the form of (clarifying) questions.

Interaction between task control and scaffolding behaviours

Prior research by Jaarsma and colleagues (2018) investigated the effects of task control on expert-novice interaction. Similarities with our results include differences in scaffolding behaviours depending on participant control. Experts cued in 83.1% of expert control segments, compared to 42.2% of novice control segments, and provided chunking in 62.9% of expert control segments compared to 13.8% of novice control segments. It is important to note that in our design, expert control always preceded novice control, meaning experts present most new information during the expert control phase by design. Novices cued in 92.2% of novice control segments compared to 26.6% of expert control segments, suggesting control still strongly influences the behaviours exhibited by both participants.

Often during instances of novice control, experts choose not to provide cueing or chunking to let novices take on responsibility, in line with findings from Van Nooijen and colleagues (2024). Instead, they would monitor novice behaviour (coinciding with high JVA) and respond to errors when prompted by the novice, sometimes through questions but most often through pauses or errors in the novices’ actions. During such pauses or errors, experts had to adjust their troubleshooting strategies from chunking to cueing to prompt actions from the novices. For instance, in the feature comparison task, the 3D car needs to be placed on the road to match the car in the security footage. It is possible to accidentally 'lift' the car up off the road in the software. When novices made this error, the experts would start with the conceptual (chunking) explanation of what went wrong, e.g. "you lifted the car up off the road", which novices struggled/were unable to convert into actions. Experts then subsequently switched to cueing, "click on the axis here", to prompt novice action. This finding sheds light on the way expertise is cognitively structured; the experts operate on a more conceptual level, and troubleshoot on what conceptually has gone wrong. This finding is reminiscent of findings from Tambaum and Nomak: "*The number of failed attempts was relatively big because during the first half of the session the tutors did only present declarative knowledge, e.g. what to do (take the cursor there). By the end of the session, both tutors of beginners redirected their focus from the display to the learner and phrased procedural knowledge e.g. how to do (how the mouse should be handled to get the cursor there)*" (Tambaum & Normak, 2018, p. 242). This result is directly translatable to practice; in contexts where experts work with novices with low conceptual understanding, starting with a hands-on practical cue when troubleshooting followed by the 'why' may lead to improved learning as opposed to starting with the 'why'.

Reciprocal gaze

Table 4

Additional Gaze Types Across Six Recorded Cases

	Both tasks	FA 1	FA 2	FA 3	FE 1	FE 2	FE 3	Total	Face task	Feature task
Reciprocal gaze	**17.5%**	7	0	2	9	5	19	42	78.6%	21.4%
Expert looks at novice only	**19.2%**	11	5	9	6	9	6	46	45.7%	54.4%
Novice looks at expert only	**0.8%**	0	0	0	0	2	0	2	100.0%	0.0%

Note. Total expert and novice gaze types (bold) represent percentages of the 240 total segments. These gaze types are additional, meaning they occur in segments where looking at the screen (high or low JVA) also occurred.

As indicated in Table 4, reciprocal gaze was present in 17.5% of segments where looking at the screen (high or low JVA) was also present, occurring most often in the face comparison task. It is important to note that this is likely an underestimation of the total reciprocal gaze present in each video, which should be considered in future research.

The most frequently occurring additional gaze type was the expert looking at the novice without the novice returning the gaze. This gaze type occurred in 19.2% of all segments, which connects to the notion posited by Kendon (1967) that eye contact from the expert may serve to gauge the novice's engagement and understanding through facial expressions.

Interestingly, novices almost never looked at the expert without the expert reciprocating the gaze; this occurred in only 0.8% of all segments, always occurring in the face comparison task. Qualitative analysis shows that these segments in the face comparison task are all instances of the expert indicating a facial feature on the side of their own face (e.g. an earlobe), thus pointing the side of their face to the novice and not reciprocating the gaze.

Participant performance

Table 5

Participant Task Performance

	FA1	FA2	FA3	
Percent agreement with expert	67%	58%	65%	
	FE1	FE2	FE3	Expert value
Standard deviation of 10 estimations	1.69m	1.59m	1.41m	1.36m

Participants completed the task they learned semi-independently during the novice control period. Scores shown in Table 5 confirm the novice level of the participants relative to the experts. Interestingly, there may be a relationship between amount of time spent and similarity to expert assessment, as in both the feature comparison and face comparison task, novices who spent more time completing the session performed better on the subsequent performance test.

Methodological novelty

In the present analysis, we chose to isolate ten moments of highest and lowest joint visual attention over time, averaged over a span of 1-10 seconds. In total, 240 10-second clips were isolated from these highest-JVA and lowest-JVA moments, and these were analysed and coded qualitatively. A limitation of this approach is that only a smaller portion of the overall corpus was qualitatively analysed to keep detailed analysis feasible. Despite this, this method provided a quantitative structure and a point of comparison within videos, allowing the qualitative conclusions that were drawn to hold more weight. This may provide a standard for 'best practice' for future research using similar approaches.

A main limitation of the novel approach presented in this research is the lower resolution gaze output that wearable eye tracking is known to have relative to its screen-based counterparts. Despite this, the use of ArUco corners with wearable eye tracking allows for several unique advantages over traditional screen-based methods of measuring JVA. Firstly, as screen-based eye-trackers can only track one pair of eyes at a time, previous JVA research required dyad members to sit at two separate screens, which impedes the participants' use of natural communication such as reciprocal gaze and gesturing. Even if these do occur, screen-based eye tracking is incapable of measuring reciprocal gaze and gesturing, for which other methods such as video recording are necessary. Having these natural communication behaviours occur within the recordings that are used for JVA analysis allows for the unique phenomenon of seeing how participant gestures towards the ArUco-framed area guide JVA in real time. We believe in the present research this rich data gain outweighs the limitation of lower resolution.

We implemented a novel method to capture JVA, together with descriptive quantitative statistics and qualitative analysis. This method is practical and portable, allowing for on-site implementation in naturalistic settings with hard-to-reach participant populations, such as our population of registered forensic experts at the Netherlands Forensic Institute. Additionally, this method has the potential to be applied on a broader scale quantitative experimental or intervention research. Although we look forward to future work (including our own) that explores quantitative research questions using this methodology, the authors recommend retaining a layer of qualitative analysis with this methodology as it allows for a 'reality-check' to see what is truly happening in the learning interactions.

Conclusions

The present research investigated the relationship between scaffolding behaviours and gaze behaviours in six recordings of learning interactions between forensic experts and novices. We found that characteristic behaviours exhibited by experts, such as providing cues and monitoring the novice's actions, coincided and appeared to result in high JVA. Given that JVA is correlated with interaction and collaboration quality, and novice learning and performance (Becker et al., 2021; Jermann et al., 2011; Schneider et al., 2015; Schneider & Bryant, 2024; Schneider & Pea, 2013; Sharma et al., 2020), understanding the types of expert and novice behaviours that may increase JVA in expert-novice dyads is of great relevance. Shvarts and Abrahamson (2019) describe JVA as the natural result of joint action; from our qualitative analysis, we consider JVA to be the measurable outcome of effective cueing behaviours from the expert or novice. When an expert points at a new tool, or a novice clicks the wrong button, JVA is the indicator that the other participant in the interaction has effectively followed that cue. This leads to information reduction in the working memory of the participant that follows, whether that be a novice seeing which tool is the correct one to use, or an expert seeing where a mistake is being made. This information reduction allows the participant that follows to act with appropriate contingency, which is central to effective scaffolding.

Our research also investigated scaffolding in the form of chunking behaviours, finding no apparent relationship between chunking behaviours and JVA. As previously described, this finding may be explained by the relationship between chunking and a different gaze behaviour; reciprocal gaze. We suggest that further research could focus on the relationship between effective expert chunking behaviours, the cognitive load reduction that follows from effective chunking, and reciprocal gaze.

The present research extends the understanding of experts and novices in one-to-one interactions by not only considering expert cueing and chunking behaviours as attempts to manage the working memory of the novice, but also by framing cueing and chunking behaviours made by novices as attempts to manage their own working memory. This is particularly prevalent in novice-controlled interactions but also visible in expert-controlled interactions. We believe further investigating the effects of cueing and chunking during one-on-one interactions on novice learning can lead to useful insights to improve the quality of learning in one-on-one interactions.

References

Becker-Genschow, S., Mukhametov, S., Pawels, P., & Kuhn, J. (October 20, 2021). Using mobile eye tracking to capture joint visual attention in collaborative experimentation. Paper presented at the https://doi.org/10.1119/perc.2021.pr.Becker

Beebe, N. L., & Guynes, J. (2006). Digital forensics curriculum development: Identification of knowledge domains learning objectives and core concepts. Twelfth Americas Conference on Information Systems (AMCIS), August 4-6, 2006, Acapulco, Mexico.

Bradski, G. (2000). The OpenCV Library. *Dr. Dobb's Journal of Software Tools*.

Dalzel-Job, S., Oberlander, J., & Smith, T. J. (2011). Don't look now: The relationship between mutual gaze, task performance and staring in second life. Paper presented at the 832-837. http://csjarchive.cogsci.rpi.edu/proceedings/2011/

Emhardt, S. N., Kok, E., Jarodzka, H. M., Brand - Gruwel, S., mm, C., & van Gog, T. (2020). How experts adapt their gaze behavior when modeling a task to novices. *Cognitive Science, 44*(9), 1-26. https://doi.org/10.1111/cogs.12893

Emhardt, S. N., Kok, E., van Gog, T., Brandt-Gruwel, S., van Marlen, T., & Jarodzka, H. (2023). Visualizing a task performer's gaze to foster observers' performance and learning—a systematic literature review on eye movement modeling examples. *Educational Psychology Review, 35*(1), 23. https://doi.org/10.1007/s10648-023-09731-7

Fiorella, L., Stull, A. T., Kuhlmann, S., & Mayer, R. E. (2019). Instructor presence in video lectures: The role of dynamic drawings, eye contact, and instructor visibility. *Journal of Educational Psychology, 111*(7), 1162–1171. https://doi.org/rg/10.1037/edu0000325

Gegenfurtner, A., Lehtinen, E., & Säljö, R. (2011). Expertise differences in the comprehension of visualizations: A meta-analysis of eye-tracking research in professional domains. *Educational Psychology Review, 23*(4), 523-552. https://doi.org/10.1007/s10648-011-9174-7

Gobet, F., Lane, P. C. R., Croker, S., Cheng, P. C., Jones, G., Oliver, I., & Pine, J. M. (2001). Chunking mechanisms in human learning. *Trends in Cognitive Sciences, 5*(6), 236-243. https://doi.org/10.1016/S1364-6613(00)01662-4

Growns, B., Dunn, J. D., Mattijssen, E. J. A. T., Quigley-McBride, A., & Towler, A. (2022). Match me if you can: Evidence for a domain-general visual comparison ability. *Psychonomic Bulletin & Review, 29*(3), 866-881. https://doi.org/10.3758/s13423-021-02044-2

Haataja, E., Garcia Moreno-Esteva, E., Salonen, V., Laine, A., Toivanen, M., & Hannula, M. S. (2019). Teacher's visual attention when scaffolding collaborative mathematical problem solving. *Teaching and Teacher Education, 86*, 102877. https://doi.org/10.1016/j.tate.2019.102877

Harper, J. M. (2012). *Mastering autodesk 3ds max 2013* (1. Aufl. ed.). Wiley.

Hessels, R. S., Teunisse, M. K., Niehorster, D. C., Nyström, M., Benjamins, J. S., Senju, A., & Hooge, I. T. C. (2023). Task-related gaze behaviour in face-to-face dyadic collaboration: Toward an interactive theory? *Visual Cognition, 31*(4), 291-313. https://doi.org/10.1080/13506285.2023.2250507

Jaarsma, T., Boshuizen, H. P. A., Jarodzka, H. M., & van Merriënboer, J. J. G. (2018). To guide or to follow?: Teaching visual problem solving at the workplace. *Advances in Health Sciences Education : Theory and Practice, 23*(5), 961-976. https://doi.org/10.1007/s10459-018-9842-1

Jarodzka, H., Scheiter, K., Gerjets, P., & van Gog, T. (2010). In the eyes of the beholder: How experts and novices interpret dynamic stimuli. *Learning and Instruction, 20*(2), 146-154. https://doi.org/10.1016/j.learninstruc.2009.02.019

Jarodzka, H., Skuballa, I., & Gruber, H. (2021). Eye-tracking in educational practice: Investigating visual perception underlying teaching and learning in the classroom. *Educational Psychology Review, 33*(1), 1-10. https://doi.org/10.1007/s10648-020-09565-7

Jermann, P., Mullins, D., Nüssli, M., & Dillenbourg, P. (2011). Collaborative gaze footprints: Correlates of interaction quality .https://doi.org/10.22318/cscl2011.184

Kendon, A. (1967). Some functions of gaze-direction in social interaction. *Acta Psychologica, 26*, 22-63. https://doi.org/https://doi.org/10.1016/0001-6918(67)90005-4

Köpsén, S., & Nyström, S. (2012). Learning in practice for becoming a professional forensic expert. *Forensic Science International, 222*(1), 208-215. https://doi.org/10.1016/j.forsciint.2012.05.026

Koskela, I., & Palukka, H. (2011). Trainer interventions as instructional strategies in air traffic control training. *Journal of Workplace Learning, 23*(5), 293-314. https://doi.org/10.1108/13665621111141902

Kroon, B., Hanjalic, A., & Boughorbel, S. (2007). Comparison of face matching techniques under pose variation. Paper presented at the *Proceedings of the 6th ACM International Conference on Image and Video Retrieval,* Amsterdam, The Netherlands. 272–279. https://doi.org/10.1145/1282280.1282323 https://doi.org/10.1145/1282280.1282323

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica, 22*(3), 276-282. https://doi.org/10.11613/bm.2012.031

Niehorster, D. C., Hessels, R. S., Benjamins, J. S., Nyström, M., & Hooge, I. T. C. (2023). GlassesValidator: A data quality tool for eye tracking glasses. *Behavior Research Methods,* https://doi.org/10.3758/s13428-023-02105-5

Nückles, M. (2021). Investigating visual perception in teaching and learning with advanced eye-tracking methodologies: Rewards and challenges of an innovative research paradigm. *Educational Psychology Review, 33*(1), 149-167. https://doi.org/10.1007/s10648-020-09567-5

Phillips, P. J., Yates, A. N., Hu, Y., Hahn, C. A., Noyes, E., Jackson, K., Cavazos, J. G., Jeckeln, G., Ranjan, R., Sankaranarayanan, S., Chen, J., Castillo, C. D., Chellappa, R., White, D., & O'Toole, A. J. (2018). Face recognition accuracy of forensic examiners, superrecognizers, and face recognition algorithms. *Proceedings of the National Academy of Sciences, 115*(24), 6171-6176. https://doi.org/10.1073/pnas.1721355115

Schneider, B. (2020). A methodology for capturing joint visual attention using mobile eye-trackers. *JoVE,* (155), e60670. https://doi.org/10.3791/60670

Schneider, B., & Bryant, T. (2024). Using mobile dual eye-tracking to capture cycles of collaboration and cooperation in co-located dyads. *Cognition and Instruction, 42*(1), 26-55. https://doi.org/10.1080/07370008.2022.2157418

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. D. (July 1, 2015). 3D tangibles facilitate joint visual attention in dyads. Paper presented at the *Cscl,* https://repository.isls.org//handle/1/403

Sharma, K., Leftheriotis, I., & Giannakos, M. (2020). Utilizing interactive surfaces to enhance learning, collaboration and engagement: Insights from learners' gaze and speech. *Sensors, 20*(7), 1964. https://doi.org/10.3390/s20071964

Shvarts, A., & Abrahamson, D. (2019). Dual-eye-tracking Vygotsky: A microgenetic account of a teaching/learning collaboration in an embodied-interaction technological tutorial for mathematics. *Learning, Culture and Social Interaction, 22*, 100316. https://doi.org/10.1016/j.lcsi.2019.05.003

Sporer, S. L., Gehrke, J., & McClure, K. A. (2022). Patrolling the borders: Experience with multiple out-group members is associated with the own-ethnicity effect. *Journal of Police and Criminal Psychology, 37*(2), 433-446. https://doi.org/10.1007/s11896-020-09387-x

Sutkin, G., MD, Littleton, E. B., PhD, & Kanter, S. L., MD. (2015). How surgical mentors teach: A classification of in vivo teaching behaviors part 2: Physical teaching guidance. *Journal of Surgical Education, 72*(2), 251-257. https://doi.org/10.1016/j.jsurg.2014.10.004

Tambaum, T., & Normak, P. (2018). Teenaged internet tutors' level of interactivity - by sharing tacit and explicit knowledge with older learners. *European Journal for Research on the Education and Learning of Adults, 9*(2), 229-248. https://doi.org/10.3384/rela.2000-7426.ojs161

Tobii. (2023a, *Tobii pro glasses 3 controller app.* Retrieved 30 April 2024, from https://www.tobii.com/products/eye-trackers/wearables/tobii-pro-glasses-3/controller-app

Tobii. (2023b, *Tobii pro glasses 3.* Retrieved 4th May 2024, from https://www.tobii.com/products/eye-trackers/wearables/tobii-pro-glasses-3

Valtakari, N. V., Hooge, I. T. C., Viktorsson, C., Nyström, P., Falck-Ytter, T., & Hessels, R. S. (2021). Eye tracking in human interaction: Possibilities and limitations. *Behavior Research Methods, 53*(4), 1592-1608. https://doi.org/10.3758/s13428-020-01517-x

van Gog, T., Jarodzka, H., Scheiter, K., Gerjets, P., & Paas, F. (2008). Guiding students' attention during example study by showing the model's eye movements. Paper presented at the *Beyond Knowledge: The Legacy Of,* 189-196.

van Gog, T., Jarodzka, H., Scheiter, K., Gerjets, P., & Paas, F. (2009). Attention guidance during example study via the model's eye movements. *Computers in Human Behavior, 25*(3), 785-791. https://doi.org/10.1016/j.chb.2009.02.007

van Nooijen, C. C. A., de Koning, B. B., Bramer, W. M., Isahakyan, A., Asoodar, M., Kok, E., van Merrienboer, J. J. G., & Paas, F. (2024). A cognitive load theory approach to understanding expert scaffolding of visual problem-solving tasks: A scoping review. *Educational Psychology Review,* https://doi.org/10.1007/s10648-024-09848-3

Walter, A. J. (2006). Surgical education for the twenty-first century: Beyond the apprentice model. *Obstetrics and Gynecology Clinics of North America, 33*(2), 233-236. https://doi.org/10.1016/j.ogc.2006.01.003

Wilkinson, C., & Evans, R. (2009). Are facial image analysis experts any better than the general public at identifying individuals from CCTV images? *Science & Justice, 49*(3), 191-196. https://doi.org/10.1016/j.scijus.2008.10.011

Appendix 1

Coding agreements for raters

Code	Definition
Expert looking at novice	Does the centre of the expert's gaze point overlap with the novice's face during the segment (yes/no)
Novice looking at expert	Does the centre of the novice's gaze point overlap with the expert's face during the segment (yes/no)
Reciprocal gaze	Occurs if the above 2 codes occur at the same time
Cueing	See flowchart in Van Nooijen and colleagues (2024)
Chunking	See flowchart in Van Nooijen and colleagues (2024)
Verbal	Does the participant produce words or speech, including non-word utterances like 'mhm' (yes/no)
Nonverbal	Does the participant use their hands or another part of the body (yes/no) OR Does the participant use their mouse on the screen (yes/no)
Findings	See findings 1-12 in Van Nooijen and colleagues (2024)

Appendix 2

Inter-rater reliability qualitative coding

	Cohen's kappa
Reciprocal gaze	1.0
Expert looks at novice (only)	0.8
Novice looks at expert (only)	0.8
Expert cueing	0.9
Expert chunking	0.8
Expert verbal	0.9
Expert nonverbal	0.9
Novice cueing	0.7
Novice chunking	0.8
Novice verbal	0.8
Novice nonverbal	0.7

Note. The minimum threshold of 0.6 was reached for all codes.