

On the Use of Gaze as a Measure for Performance in a Visual Exploration Task

Catharine Oertel¹, Alessia Coppi², Jennifer K. Olsen¹, Alberto Cattaneo², and Pierre Dillenbourg¹

¹ EPFL, Lausanne, Switzerland {catharine.oertel, jennifer.olsen, pierre.dillenbourg}@epfl.ch

² SFIVET, Lugano, Switzerland {alessia.coppi, alberto.cattaneo}@iuffp.ch

Abstract. Visual exploration skill acquisition is important for many vocational professions, yet many apprentices struggle to acquire these skills, impacting both their grades and practical work. Traditionally, the learning of visual skills is facilitated through exercises where it can be difficult to identify struggling apprentices early. We propose the use of gaze patterns to identify apprentices who may need additional support. In this paper, we investigated differences in gaze patterns between teachers and apprentices and the relationship between gaze patterns and student performance. In a study with 18 fashion design apprentices and 16 fashion design teacher, we found teachers have a higher gaze coverage of an image than apprentices and there is a correlation between the verbalisation score of apprentices and their similarity of gaze patterns to the teacher average. Using these results, we may be able to adapt exercises to apprentices' particular needs early in the learning process.

Keywords: Eye tracking · vocational education · visual skill acquisition.

1 Introduction

Across most domains, apprentices must learn the important skill of how to read and interpret different visualizations. Within some domains, like statistics, the visualizations are an important tool for communicating information. In other domains, such as those within vocational education, the visualizations are a core aspect to the everyday tasks in which domain professionals engage. For example, gardeners must be able to read landscaping plans, carpenters construction plans, and fashion designers fashion sketches. However, as apprentices are learning, they often struggle to explore these visualizations, and it can be difficult for the teacher to understand how to provide support given the limited amount of information that they have regarding what the apprentices are doing during the task. Currently, teachers may ask the apprentices to verbalise what they are seeing, but this limits the teachers to only being able to help a few apprentices at a time and they may have trouble identifying who needs support. In this paper, we propose using gaze patterns to identify apprentices who are struggling and identify when support is needed.

Vocational education is an important pillar in the Swiss-educational system. About 70% of all high-school apprentices choose vocational training over a university education. Vocational professions encompass many different professions (e.g., bakers, carpenter, florists and fashion designers). Across every profession, understanding visual information is an important skill that is used on a daily basis within their professional careers. However, it can be difficult for apprentices to identify and describe the relevant areas of an image even after they spend time looking at them during a task.

In the current paper, we focus on visual skill acquisition in fashion designers. A task that is typical for fashion design apprentices is to explore a photograph of a piece of clothing, for example a shirt, and to encode and identify the important details of that clothing item. Because the task involves the exploration of images, there is not a prescribed method that is correct for engaging in the task. Therefore, to identify apprentices that may be struggling, we propose to analyze their gaze patterns and how they may relate to those of experts. Specifically, we ask the following questions: First, what are the differences, if any, between apprentices and experts in a visual exploration task essential for the training of fashion designers? Second, are apprentices with a higher verbalisation score more similar to the experts in terms of gaze patterns than apprentices with lower scores?

2 Background

Within educational research, eye-tracking data has become an important source of process data. It has been used across a range of contexts to understand the learning processes that apprentices engage in [11, 12, 9]. Additionally, it has been used to provide apprentices with real-time interventions that can adapt to the apprentices' needs [13, 2]. However, much of this work has focused on providing support to problem-solving tasks and less focus has been given to image assessment, as is more common in professional domains.

On the other hand, within professional domains, eye-tracking has also been extensively used to understand how apprentices process different visual stimuli and how this may differ from experts [7, 6, 20, 4]. Across studies and domains, it was found that experts tend to have a longer fixation on the areas of importance in the images [7, 20, 19, 10] while novices spent more time looking at irrelevant areas of the images [5]. Additionally, experts compared to novices are faster at detecting the areas of interest in the images [20, 5]. Despite the longer time spent on the areas of interest, studies have also shown that experts spend more time exploring the image before focusing on an area [6, 4] and that the timing of a behavior may matter [7]. In other words, experts are able to identify the important areas of the image and spend more time observing the relevant features and less time on irrelevant ones. These differences can even be found between individuals with small differences in expertise [18] and seen within the same individual as they gain expertise over time [3].

However, these differences may be mitigated by the task. For example, professionals exposed to a non-professional task (e.g., search task) will not necessarily differ from non-experts [8] and tasks that are simple have not shown a discrimination between novices and experts [17]. Many of the above findings have been done with identification tasks in which the goal of the task is to find a mistake in the image. For example, in radiology, the novices need to detect if there is a fracture visible in the image [20]. In these cases, the task is a search task rather than an exploration task, as is the case for fashion designers.

Additionally, much of the work with gaze on images has focused on areas of interest (AOIs), which treats all other areas as irrelevant. More recent work in education has incorporated a grid rather than AOIs as to treat all areas of the screen as equally important [15, 9]. Using a grid, we are able to discover places that may be of interest on the image that were not identified ahead of time.

2.1 Contributions

In the current paper, we contribute to the literature on gaze analysis by providing further insights into the relationship between novice and expert gaze patterns and their relationship to performance. Additionally, we contribute to learning on visual exploration tasks by furthering our understanding of how apprentices engage with the image in a specific context, namely with fashion designers. In contrast to previous research, we focus on apprentices engaged in a visual exploration task, which is very typical in vocational education, rather than identifying mistakes [7] or problematic areas [6], [20]. Similar to [5], we focus on apprentices identifying characteristics of the visualization. However, unlike [5], we focus on static images rather than dynamic stimuli. Additionally, we use a methodology that allows for calculating the amount of similarity between teachers and apprentices without the need to predefine AOIs, which are typically used in learning with visualisations.

3 Methods

The data presented in this paper is based on the data gathered within the context a larger study described in [1]. To investigate gaze patterns between apprentices and teachers in a visual exploration task we focused on a typical task for second year apprentices in a three-year program. The task consisted of the visual exploration of shirts described in more detail below.

3.1 Participants

A total of 34 participants from two vocational fashion design schools in Switzerland participated in our study. There were 18 fashion design apprentices and 16 fashion design teachers. The apprentices were all in their second year of a three-year certificate program. They had little or no experience with sewing shirts. On the other hand, all the teachers were regarded as experts. They completed

fashion design training, had 10-20 years of working experience and taught at the fashion design school.

3.2 Procedure

For the study, we used a pull-out design that took place at the apprentices' schools but where the participants were asked to enter a separate room from their normal classroom. In this room, a computer with an external screen and an eye-tracker was set up. Specifically, we used a Tobii Pro X2-60 eyetracker and a 15 inch PC screen.

At the beginning of each session, the participants were provided with basic information about the experiment and the functioning of the eyetracker. The participants were asked to avoid movements such as tilting or turning the head, looking away from the screen or at the researcher and obstructing the eyetracker in any way. After the instructions, the eyetracker was calibrated for each participant.

For primary task, the participants were asked to observe a set of five images, which were presented to participants on a screen. Each of the images showed a model wearing a shirt. Each of the shirts was of a different cut and featured specific details such as for example embroidery or special buttons. The images are described in further detail in the following section. Apprentices were told to observe the details of the image and to verbally describe them. Each of the images was presented to participants for 40 seconds. Between each image there was a black screen for 20 seconds.

3.3 Stimuli

The images used in the study were identified with the help of the teachers. They were chosen to resemble as closely as possible the types of images used typically during fashion design training. Specifically, the teachers proposed the use of photographs representing models wearing different kind of shirts with a range of peculiarities related to shape, style, details and parts. Different parts in a shirt comprise the middle part, the shoulders, sleeves, opening and hem. Teachers who were not participating in this study, provided detailed descriptions of the images. These descriptions included information on the details of the shirt such as the opening (an open section at the top of a garment for the neck of the wearer), shoulders, stitching, pockets, sleeves, cuffs, hem and buttons. All of these parts are essential for apprentices to pay attention to and to visually explore. An example of an image used can be seen in Fig 3.3.

3.4 Variables

To quantify gaze patterns without the need to manually assign areas of interests and find a more global measures applicable to any picture, we defined and calculated a similarity measure between gaze distributions as well as calculated the gaze coverage of the image and the fixations on the image. As a ground truth for apprentices' performance, we used their verbalisation score.



Fig. 1. Example stimuli focusing on a shirt design.

Similarity Calculation The gaze similarity is a measure of how much an individual is looking at the same region of the image as an other individual during an interval of time. To quantify the attention towards regions, we computed the main region of interest of an image by finding the smallest rectangular area containing all the fixations of all participants for the given image. Given the size of the remaining region of interest is then divided into a regular grid of size 16x10. Afterwards for a given participant and time window, we counted the amount of gaze hits towards each cell of the grid. This process led to a vector representation of 160 dimensions containing the per-region gaze counts, which we will refer as the *gaze count vector*. This gaze count vector was used to compute both our similarity and gaze coverage measures.

To compute the similarity between two individual gaze count vectors, we normalized the vectors to be unitary and computed their scalar product as in [14]. Specifically, to compute the similarity between an apprentice and the average of all teachers, we computed gaze count vectors over time windows of 5 seconds for the entire 40 seconds for the apprentice and each teacher. We then created a prototypical teacher by averaging the gaze count vectors over all teachers within the same time windows. Using the prototypical teacher, we computed the similarity with the apprentice for every time window and computed an average.

Gaze Coverage Calculation We define gaze coverage as the proportion of the image which is observed by the individual at least once. To compute the coverage, we first retrieved the heatmap for the given main region of interest, estimated over the duration of the stimuli. Using this heatmap, we computed the gaze coverage as the ratio of pixels which were above a threshold based on a fixed value over the total amount of pixels in the image. In our experiments, the

threshold was set to 0.005, given that the heatmap values were normalized such that the maximum value was 1.

Verbalisation Score The verbalisation score is calculated for each image and for each apprentices. Specifically that means that teachers would listen to an audio recording of apprentices’ verbalisations. They would count the number of uniquely correctly identified and named details and divide this by the number of actual items to be identified times 100.

4 Research Questions

In order to answer the research questions defined in the introduction section of this paper we define the following sub-questions. Both R1 and R2 are designed to further define the question of whether there are quantifiable differences between the gaze patterns of apprentices and teachers. R3 clarifies the question whether apprentices with a higher verbalisation score are more similar to the teachers in terms of gaze patterns.

R1: Are there significant differences in terms of gaze exploration between apprentices and teachers ? In previous studies it has been found that experts and apprentices differ in the amount to which they explore the image/video[6, 4]. In order to answer this question we calculated the amount of “gaze coverage” for all images and compare the groups of apprentices and teachers.

R2: Are there significant differences between apprentices and teacher in terms of fixation on area of interests? As discussed in the background section, differences in terms of fixations on areas of interests have been found in studies carried out on related by different tasks [7, 20, 19, 10]. Therefore, in the current study, we investigate whether we can also find significant differences in fixation for this visual exploration task.

R3: Are apprentices with a higher verbalisation score more similar to the teachers in terms of gaze patterns? In order to answer this research question, we calculated the similarity of gaze distribution patterns between the average teacher (the average of gaze distribution patterns of all teachers) and each participant. We then correlated each apprentices’ verbalisation score with their similarity score.

5 Results

R1: To test if there were differences in gaze exploration between apprentices and teachers, we ran an ANOVA comparing the two groups in terms of their gaze coverage of the image. A boxplot illustrating the proportionate amount of gaze coverage of the image between the groups of teachers and the group of apprentices can be seen in Fig. 5. The ANOVA tests revealed that the group category had an effect on the gaze coverage $F(1,163) = 11.42$, ($p < .01$) with the teachers having more gaze coverage than the apprentices.

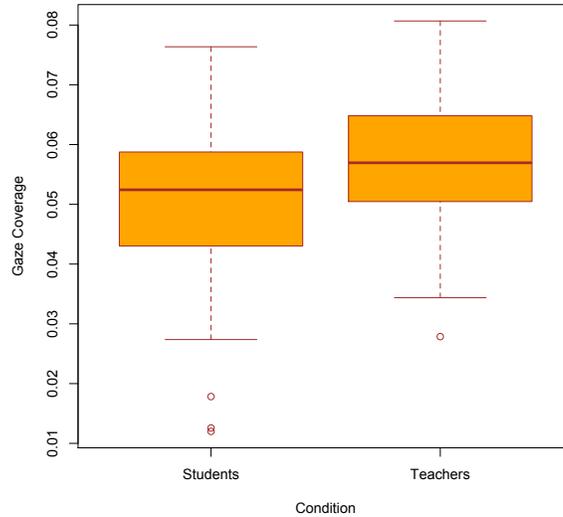


Fig. 2. Gaze Coverage contrasting teachers and apprentices.

R2: To test if there were differences in the time spent on the predefined important area of interests in each picture, we conducted a MANOVA since there is dependence between the time spent in each area. The MANOVA examined the average fixation per participant in the areas of interest as dependent variables with the teacher and apprentice classification as the independent variable. No significant multivariate effect could be found, $F(10,240)=1.38$ ($p=.25$).

R3: Finally, a Pearson correlation was computed to assess the relationship between apprentices' verbalisation score and their gaze similarity to the average teacher. One participant was removed before analysis due to missing data. There was a positive correlation between the two variables, $r(15) = .87$ ($p<.01$), with the higher the verbalisation score the more similar the gaze pattern.

6 Discussion

In this paper we investigated gaze patterns between apprentices and teachers in an image exploration task that is relevant to vocational education training for fashion designers. We found differences of gaze patterns distinguishing teachers and apprentices. Specifically, we found that teachers have a higher gaze coverage of the image than apprentices. This result is in line with previous research where it was found that experts spend more time exploring the image [6, 4]. Specifically in our study, this result may have been due to the nature of our task in which

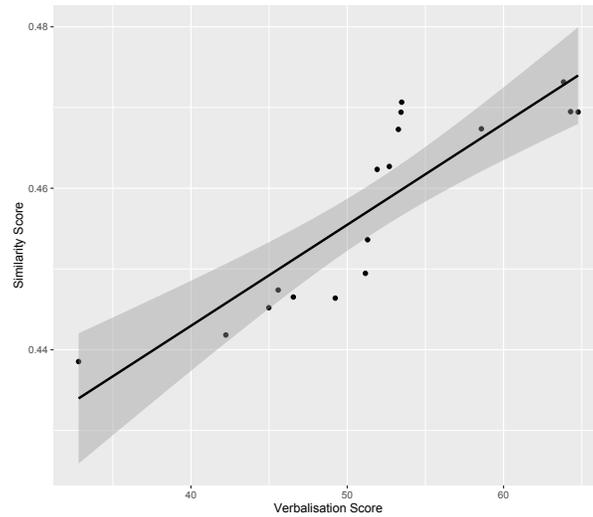


Fig. 3. Scatter plot illustrating the relation between verbalisation and similarity score.

the goal was to explore (i.e. spend time looking at more of the image) where this may have been an acquired skill by the teachers.

Further, in this study, we did not find any significant differences between the teachers and apprentices with the amount of time spent on the important areas of interest. A factor that might have contributed to this finding is that in the case of this experiment, the images the clothing design teachers recommended as stimuli, displayed several areas that were of relevance. These were also the same areas that were the most visually prominent ones and would attract the visual attention of complete novices. This prominence might have contributed to there not being any clear differences in proportionate amount of fixations between experts and apprentices.

We also found a correlation between apprentices' verbalisation score and similarity of gaze patterns to the average teacher. The higher the verbalisation score the more similar the gaze pattern. This finding is in line with previous research indicating that there are also differences in gaze patterns where the expertise level is small between novices and experts [18]. For future work, it might be interesting to compare apprentices' gaze patterns not to the average teacher but to the teacher who achieves the highest verbalisation score. While using an average teacher has the advantage that not everything is modeled based on the gaze behaviour of one expert alone and takes into account different ways of exploring the images, it also bears the danger that differences are being averaged out.

One limitation of the study is the duration to which apprentices were exposed to the stimuli. Forty seconds is in hindsight a rather long time for exploring the

stimulus and differences might have been more pronounced if a shorter period of time was chosen.

Finally, the aim of this study was also to work towards building a system that is able to detect if apprentices are diverging from teachers gaze patterns. Both “Gaze Coverage” and “Similarity Score” appear to be useful measures to identify such divergences. In future work, we are planning to explore the use of an embodied artificial agent to facilitate such an intervention. An artificial agent could establish joint attention through pointing gestures or also verbally indicate different areas of interest. It has been shown that artificial agents can be quite useful for creating joint attention and disambiguating confusing attention targets [16]. Moreover, the agent could summarize how long an expert/or group of experts spend on a given area and might hint at the fact that this area appears to be particularly difficult and might deserve further attention.

7 Conclusion

We found differences in gaze patterns between teachers and fashion design apprentices in a visual exploration task. In future work, we are planning to investigate how we can use these findings to support apprentices in their visual skill acquisition. One method of intervention we are currently considering is the use of cueing. By using cues, such as fashion design specific annotations, apprentices’ attention could be guided towards all relevant details of the garment.

References

1. Coppi, A., Cattaneo, A., Dillenbourg, P.: Observing like an expert: Effects of using annotations on apprentices’ gaze patterns and verbalizations. In: 6th congress on Research in Vocational Education and Training of the Swiss Federal Institute for Vocational Education and Training (SFIVET) (2019)
2. D’Angelo, S., Begel, A.: Improving communication between pair programmers using shared gaze awareness. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. pp. 6245–6290. ACM (2017)
3. Haider, H., Frensch, P.A.: Eye movement during skill acquisition: More evidence for the information-reduction hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **25**(1), 172 (1999)
4. Jaarsma, T., Jarodzka, H., Nap, M., van Merriënboer, J.J., Boshuizen, H.P.: Expertise under the microscope: processing histopathological slides. *Medical education* **48**(3), 292–300 (2014)
5. Jarodzka, H., Scheiter, K., Gerjets, P., Van Gog, T.: In the eyes of the beholder: How experts and novices interpret dynamic stimuli. *Learning and Instruction* **20**(2), 146–154 (2010)
6. Krupinski, E.A., Tillack, A.A., Richter, L., Henderson, J.T., Bhattacharyya, A.K., Scott, K.M., Graham, A.R., Descour, M.R., Davis, J.R., Weinstein, R.S.: Eye-movement study and human performance using telepathology virtual slides. implications for medical education and differences with experience. *Human pathology* **37**(12), 1543–1556 (2006)

7. Moreno, F., Reina, R., Luis, V., Sabido, R.: Visual search strategies in experienced and inexperienced gymnastic coaches. *Perceptual and Motor skills* **95**(3), 901–902 (2002)
8. Nodine, C.F., Krupinski, E.A.: Perceptual skill, radiology expertise, and visual test performance with nina and waldo. *Academic radiology* **5**(9), 603–612 (1998)
9. Olsen, J., Sharma, K., Alevin, V., Rummel, N.: Combining gaze, dialogue, and action from a collaborative intelligent tutoring system to inform student learning processes. *International Society of the Learning Sciences, Inc.[ISLS]*. (2018)
10. Pappas, I., Sharma, K., Mikalef, P., Giannakos, M.: A comparison of gaze behavior of experts and novices to explain website visual appeal (2018)
11. Prieto, L.P., Sharma, K., Dillenbourg, P.: Studying teacher orchestration load in technology-enhanced classrooms. In: *Design for Teaching and Learning in a Networked World*, pp. 268–281. Springer (2015)
12. Raca, M., Dillenbourg, P.: System for assessing classroom attention. In: *Proceedings of the Third International Conference on Learning Analytics and Knowledge*. pp. 265–269. ACM (2013)
13. Sharma, K., Alavi, H.S., Jermann, P., Dillenbourg, P.: A gaze-based learning analytics model: in-video visual feedback to improve learner’s attention in moocs. In: *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge*. pp. 417–421. ACM (2016)
14. Sharma, K., Caballero, D., Verma, H., Jermann, P., Dillenbourg, P.: Looking at versus looking through: A dual eye-tracking study in mooc context. *International Society of the Learning Sciences, Inc.[ISLS]*. (2015)
15. Sharma, K., Olsen, J.K., Alevin, V., Rummel, N.: Exploring causality within collaborative problem solving using eye-tracking. In: *European Conference on Technology Enhanced Learning*. pp. 412–426. Springer (2018)
16. Skantze, G., Hjalmarsson, A., Oertel, C.: Turn-taking, feedback and joint attention in situated human–robot interaction. *Speech Communication* **65**, 50–66 (2014)
17. Tsuchiya, R., Shimizu, M., Mikami, K., Aizu, K., Hirakawa, Y., Takase, S., Nishizawa, Y.: The characteristic of nurses eye movements during observation of patients with disturbed consciousness. *Open Journal of Nursing* **7**(12), 1502 (2017)
18. Van Gog, T., Paas, F., Van Merriënboer, J.J.: Uncovering expertise-related differences in troubleshooting performance: combining eye movement and concurrent verbal protocol data. *Applied Cognitive Psychology* **19**(2), 205–221 (2005)
19. Wolff, C.E., Jarodzka, H., van den Bogert, N., Boshuizen, H.P.: Teacher vision: expert and novice teachers perception of problematic classroom management scenes. *Instructional Science* **44**(3), 243–265 (2016)
20. Wood, G., Knapp, K.M., Rock, B., Cousens, C., Roobottom, C., Wilson, M.R.: Visual expertise in detecting and diagnosing skeletal fractures. *Skeletal radiology* **42**(2), 165–172 (2013)