Journal of Expertise
2023. Vol. 6(3) © 2023.

ISSN 2573-2773

Measuring Teachers' Visual Expertise Using the Gaze Relational Index Based on Real-World Eyetracking Data and Varying Velocity Thresholds

Christian Kosel*[1], Angelina Voggenreiter*[2], Jürgen Pfeffer[2], and Tina Seidel[1]

[1] Department of Educational Sciences, Technical University Munich, Germany
[2] Department of Computational Social Science and Big Data, Technical University Munich, Germany
*These authors contributed equally.

Correspondence: Christian Kosel, christian.kosel@tum.de

Abstract

This article contributes to our understanding of teachers' visual expertise by measuring visual information processing in real-world classrooms (mobile eye-tracking) with the newly introduced Gaze Relational Index (GRI) metric, which is defined as the ratio of mean fixation duration to mean fixation number. In addition, the aim of this article was to provide a methodological contribution to future research by showing the extent to which the selected configurations (i.e., varying velocity thresholds and fixation merging) of the eye-movement-event-detection algorithm for detecting fixations and saccades influence the results of eye-tracking studies. Our study has two important take-home messages: First, by adopting a novice-expert paradigm (two novice teachers and two experienced teachers), we found that the GRI can serve as a sensitive measure of visual expertise. As hypothesized, experienced teachers' GRI was lower than that of the novice teachers, suggesting that their more fine-graded organization of domain-specific knowledge allows them to fixate more rapidly and frequently in the classroom. Second, we found that the selected velocity threshold parameter alters and, in the worst-case scenario, biases the results of an eye-tracking study. Therefore, in the interest of the further generalizability of the results within visual-expertise research, we emphasize that it is highly important to report configurations that are relevant to the identification of eye movements.

Keywords

Visual expertise, information processing, eye tracking, fixations, professional vision

Introduction

Human visual expertise in vision-intensive domains, such as medicine (Gegenfurtner & Seppänen, 2013), sports (Aglioti et al., 2008), driving (Lappi et al., 2017), and aviation (Peißl et al., 2019), reflects complex cognitive and visual processing that has been developed by domain experts through deliberate and consistent practice over a long period (Gegenfurtner et al., 2011). Domain experts, in comparison to domain novices, are more skilled in the perception, interpretation, and evaluation of domain-specific visual information (Gegenfurtner, 2020). In recent years, there has been growing interest in the visual expertise of teachers because they must perceive and interpret a large amount of dynamic visual information to effectively manage the complexity of a classroom full of students. A rich body of studies has revealed that experienced and novice teachers differ markedly in their visual processing (see meta-analysis/reviews in Gegenfurtner et al., 2011; Grub et al., 2020).

In addition to studies based on verbal reports and think-aloud protocols that were collected after a visual stimulus (i.e., video vignette, photograph; van Es & Sherin, 2010) was shown to the participating teachers, other studies explored teachers' visual expertise based on fine-graded data that were collected with eye-tracking devices (Kosel et al., 2023; McIntyre & Foulsham, 2018). Eye-tracking is an effective method with which to explore where, how often, and how long teachers direct their visual attention (Holmqvist et al., 2015). In visual expertise research across various disciplines, including teaching, two eye-tracking parameters have been found to be sensitive to expertise: the number of fixations and average fixation duration (Gegenfurtner et al., 2011; Grub et al., 2020).

Experts tend to exhibit more fixations on areas relevant to their profession, which is related to their ability to extract relevant information efficiently from these areas, suggesting top-down knowledge and focused visual attention. Novices, on the other hand, may exhibit fewer fixations because they must spread their attention across multiple areas to gather the necessary information, without having the ability to assess the relevance of each area quickly (Grub et al., 2020).

Average fixation duration provides another insight into the efficiency and depth of visual processing. Experts often have shorter average fixation durations than novices. Researchers have linked experts' faster encoding of information to their deeper domain knowledge, which allows them to efficiently process and represent information within their domain. As a result, experts can plan ahead and make informed decisions, especially in dynamic situations in which quick thinking and adaptation are essential (Chi & Glaser, 1988).

These findings related to the number of fixations and average fixation duration underscore the fact that the processing of visual information varies not only across individuals but also across levels of expertise (Gegenfurtner et al., 2011). Until now, however, research on visual expertise has lacked a single common metric via which to capture and contrast the visual expertise of domain experts and novices. Lowe and Boucheix (2016) fill this gap by introducing a novel eye-tracking metric called the Gaze Relational Index (GRI) that is indicative of expert visual processing.

The GRI is defined as the ratio of mean fixation duration to mean fixation count. As far as we know, the GRI has only been applied in two studies that used data derived from laboratory-based stationary eye-movement recordings. Gegenfurtner and colleagues investigated the GRI regarding 3D dynamic medical visualizations in diagnostic radiology (Gegenfurtner et al., 2020), and Grub and colleagues focused on experienced and novice teachers' gazes using video vignettes (Grub et al., 2022). These studies found either marginal expertise differences (Gegenfurtner et al., 2020) or no differences (Grub et al., 2022).

The present study aims to go beyond eye-tracking data collected in the laboratory. In laboratory-based approaches, researchers collect on-action data, in which participants' eye-tracking data are collected during the passive viewing of a pre-recorded stimulus, such as a video sequence or series of photos. However, this study is intended to apply the GRI to the study of teachers' visual expertise by calculating GRI values based on experienced and novice teachers' real-world eye-gaze data, which will be extracted from a mobile eye-tracking device. This in-action data refers specifically to the eye-movement data that are collected while the participants are actively engaged in real-world tasks or activities.

Another challenge that arises in the domain of visual expertise is the fact that the majority of eye-tracking-based studies do not implement, in their reports, information about how the various eye-tracking parameters are calculated; for example, which specific configurations in the eye-movement-event-detection algorithm were used to detect fixations and saccades. However, this is especially important, as an increasing number of researchers are analyzing their gaze data via advanced external analysis tools and scripts (Dolezalova & Popelka, 2016; Panetta et al., 2020) based on the raw eye-tracking data extracted from an eye tracker. We, therefore, aim to demonstrate the extent to which the use of different configurations may affect the detection of fixations and saccades and, thus, also the interpretation of the results. The above-outlined GRI is a suitable measure via which to investigate how different configurations affect the results of eye-movement experiments, as it is a single-value measure that allows straightforward comparisons.

Professional Vision and Visual Expertise of Teachers

Professional vision is commonly used as a conceptual framework in the field of cognitive-oriented teacher research (Goodwin, 1994; Seidel et al., 2014). The concept implies a two-step-process: (1) noticing, which describes teachers' ability to selectively direct their attention to relevant events in the classroom, and (2) knowledge-based reasoning, which refers to teachers' ability to interpret these events based on their professional knowledge (Seidel et al., 2014; van Es & Sherin, 2010). Thereby, noticing and knowledge-based reasoning are not isolated processes; rather, they interact with one another (Seidel et al., 2014). Teachers' professional knowledge drives their attentional processes in a top-down manner (i.e., selective attention is inferred from their knowledge), and in turn, noticing activates teachers' knowledge in the form of curriculum scripts and classroom routines that are stored in their long-term memory (i.e., teachers can make sense of what they see; Lachner et al., 2016). This implies that professional vision is formed primarily through consistent practice over many years, during which teachers accumulate professional knowledge, as well as that professional vision is primarily a characteristic of experienced teachers (Berliner, 2001).

The concepts of professional vision and visual expertise in educational research refer to the ways in which teachers perceive and interpret visual information in their professional practice. However, visual expertise focuses specifically on the development of advanced visual skills and knowledge in a particular domain. It refers to the ability of teachers to accurately and efficiently perceive, analyze, and interpret visual information related to their area of expertise. Visual expertise involves becoming highly skilled at recognizing and understanding visual patterns, cues, and indicators that are relevant to your subject matter, pedagogical strategies, or student needs (Gegenfurtner et al., 2011; Gegenfurtner et al., 2022).

Eye-tracking studies performed at the intersection of professional vision and visual expertise provide further evidence that teachers' visual processes change as their expertise levels increase (Grub et al., 2020; Kosel et al., 2023; van den Bogert et al., 2014). For example, it has been found that experienced teachers, as compared to novice teachers, are able to distribute their attention more evenly across students (van den Bogert et al., 2014) and monitor a larger group of students during teaching (Kosel et al., 2021; 2023). Beyond these findings, which are especially relevant to classroom management, studies have shown that experienced teachers, like domain experts in other vision-intensive fields (Gegenfurtner et al., 2011; Gegenfurtner et al., 2022), have shorter but more fixations, while domain novices have longer but fewer fixations. Because it is assumed that fixations indicate that information is being perceived and processed cognitively (Rayner, 2009), the results suggest that experts encode information more rapidly because of their more advanced and fine-grained knowledge structures, which drive visual attention in a top-down manner (Gegenfurtner et al., 2022).

In contrast, novices do not have this accumulated knowledge, and their attention is more affected by the external and salient features of the visual stimulus in a bottom-up manner (Gegenfurtner et al., 2022). The rapid information processing of experienced teachers, as reflected in their short fixation durations, is also consistent with Ericsson and Kintsch's (1995) theory of long-term working memory. They state that experts increase the capacity of their working memory by building retrieval structures in their long-term memory (see also Gegenfurtner et al., 2022). The knowledge embedded in this retrieval structure is available in the working memory and enables experts to process visual information more rapidly than novices, who have not yet fully developed a knowledge-based retrieval structure. In other words, not only having a large amount of domain-relevant knowledge but also having a superior organization of this knowledge are relevant to visual expertise.

In addition, the ability of experts to process more information, as reflected in higher numbers of fixations, is related to the assumptions of Haider and Frensch's (1996) information-reduction hypothesis. They argue that experts optimize the amount of information processed by separating task-relevant from task-irrelevant information. Ignoring redundant information leads to experts having more capacity in their working memory to use in processing relevant information. Both theories are also important parts of the Cognitive Theory of Visual Expertise (CTVE; Gegenfurtner et al., 2022), which covers additional important aspects of visual expertise (e.g., parafoveal and holistic information processing). Taken together, the outlined assumptions help us understand experienced teachers' faster and more automated information processing as involving less conscious effort, suggesting that experienced teachers encode and update dynamic teaching situations, which involve many short fixations, more rapidly than novices (Gegenfurtner et al., 2020; Grub et al., 2020).

To capture visual expertise using a single expertise-sensitive metric, Lowe and Boucheix (2016) developed and introduced the GRI. The GRI is an empirical measure of the relative extent to which various areas of interest were attended to during the visual exploration of a stimulus. It is directly based on eye-tracking data and defined as the ratio of mean fixation duration to mean fixation count (see also Gegenfurtner et al., 2020). The GRI can be used to provide important profiles across the various components of a display, with differences in these profiles between individuals being interpreted as an indicator of their levels of domain-specific expertise. Based on previous empirical studies (e.g., Wolff et al., 2016) and the way in which the GRI is calculated, it can be inferred that the GRI should be higher for novice teachers than experienced teachers.

Because the GRI is still emerging in the field of visual expertise, the number of studies to date is limited. In a study by Gegenfurtner and colleagues (2020), the GRI was calculated for dynamic 3D medical visualizations. They found that the GRI was slightly but statistically non-significantly higher for novices as compared to experts (Gegenfurtner et al., 2020). In the educational context, Grub and colleagues (2022) analyzed the GRI in a standardized experimental design in which experienced and novice teachers perceived various classroom situations via short video sequences.

Contrary to their hypothesis, they found no differences in the number and duration of fixations and, thus, no differences in the GRI (Grub et al., 2022). However, as discussed by Gegenfurtner (2020), the full potential of the GRI may come to light when an experiment is situated outside the lab, that is, when using mobile eye-tracking "to mirror the full complexity of visual input that experts routinely deal with in their everyday work surroundings" (Gegenfurtner et al., 2020, p. 38). However, the number of studies that have analyzed mobile eye-tracking data to explore expertise differences concerning the number and duration of fixations (the basis for the GRI) is limited (Huang et al., 2021).

While the in-action study performed by Huang and colleagues (2021) confirms expected expertise differences regarding the two metrics, the findings of on-action eye-tracking studies are more heterogeneous (Grub et al., 2022; Kosel et al., 2021; van den Bogert et al., 2014; Wolff et al., 2016). One reason for this, as Gegenfurtner and colleagues (2020) describe, could be that eye-tracking experiments performed in the laboratory cannot capture the full dynamic complexity that can be recorded via mobile eye-tracking in teachers' natural work environments. Thus, there is a need to further explore novice and experienced teachers' visual expertise, as measured with the GRI, using mobile eye-tracking data.

Classify Eye Movements Using Event-Detection Algorithms

Across all academic disciplines, eye-tracking-based studies rely on eye-movement-event-detection algorithms to analyze raw data and classify different types of eye movement, such as fixations (moments when the eye is relatively still and visual information is processed) and saccades (rapid eye movements between two or more phases of fixation). Many different algorithms are available today (for a review and evaluation of different algorithms, see Andersson et al., 2017). Event-detection algorithms can be broadly grouped into dispersion- and velocity-based algorithms (Andersson et al., 2017).

One of the velocity-based algorithms that is most frequently used for detecting fixations is the Identification by Velocity Threshold (I-VT) algorithm. This algorithm uses only one parameter, the fixed velocity threshold for saccade detection, where "fixations are segments of samples with point-to-point velocities below the set velocity threshold, and saccades are segments of the sample with velocities above this threshold" (Andersson et al., 2017, p. 618). The fixed and *a-priori*-defined velocity is typically given in visual degrees per second (°/s). Commonly used values for the velocity threshold in lab-based eye-tracking studies range between 5 and 50°/s, with lower values being used for oculomotor studies and higher values being used for cognitive studies (Andersson et al., 2017). The I-VT algorithm is implemented in most of the recent commercial eye-tracking software, such as Tobii Pro (Tobii, 2022). However, because fixation is a fundamental parameter of most eye-tracking studies, outcomes depend on not only the algorithm used to separate fixations from saccades (Salvucci & Goldberg, 2000) but also the various velocity thresholds employed for the algorithms (Andersson et al., 2017; Holmqvist et al., 2015). In other words, different velocity thresholds may produce significantly different results (Salvucci & Goldberg, 2000). The various velocity thresholds can be easily changed in most software solutions.

In Tobii Pro, for example, velocity thresholds of 30°/s and 100°/s are pre-stored (30°/s = fixation filter; 100°/s attention filter). In this context, Hossain and Miléus (2016) compared different velocity thresholds for fixation identification in low-sample-rate mobile eye-trackers, such as the Tobii Pro Glasses 2. They point out that the IV-T fixation filter does not perform as well on mobile eye-trackers as it does on lab-based eye-trackers, especially when many head movements are involved in the recordings. The problem is that head movements have an effect on velocities, and many fixations will not be detected by the IV-T algorithm in such cases.

They found that the default setting of 30°/s underestimates the periods during which a participant gathers information because a large proportion of smooth pursuits (eye movements in which the eyes remain fixated on a moving object) and vestibular-ocular reflex (VOR; stabilizing eye movements in the opposite direction of head movements) are classified as saccades. One way to counter this is to increase the velocity threshold of the mobile eye-tracker. Using the 100°/s attention filter would overestimate information gathering because fixations, smooth pursuits, VOR periods, and 10–15% of short saccades will be classified as fixations (Hossain & Miléus, 2016). However, Hossain and Miléus (2016) found the highest fixation-detection precision in mobile eye-tracking using a velocity threshold between 90°/s and 100°/s when head movements are involved and not compensated for with external gyroscope data.

Overall, because some studies and technical reports indicate that results are significantly changed when using different velocity thresholds (Hossain & Miléus, 2016; Olsen, 2012; Salvucci & Goldberg, 2000), the selected velocity threshold must be reported in research studies to make the results comparable. However, most studies in the educational context (Chaudhuri et al., 2021; Cortina et al., 2015) and other fields in which mobile eye tracking is used, such as aviation (Weibel et al., 2012), do not specify the velocity thresholds used to detect fixations and saccades.

In addition to velocity thresholds, so-called fixation merging is another configuration of the IV-T algorithm that must be addressed. The basic idea of merging fixations is that very short fixations (i.e., short fixations do not reflect cognitive processing) are merged with the next longer fixations in their vicinity (within 0.5° of the visual angle; Tobii, 2022). Merging can be set automatically in the Tobii software package (Olsen, 2012; Tobii, 2022). However, fixation merging has consequences regarding the classification of the number of fixations and, thus, the results obtained using fixation-based metrics, such as the GRI. However, the extent of this effect has not yet been described, which makes an assessment using various velocity thresholds relevant to future eye-tracking studies in the context of visual expertise.

The Present Study

In the present study, we aimed to explore teachers' visual expertise by adopting an expert-novice paradigm. Established expertise theories and prior empirical findings indicate that teachers, through deliberate practice over a long period, develop visual expertise, which leads to qualitatively enriched and superior ways of visually perceiving and processing information as compared to novices. Two expertise-sensitive eye-tracking metrics are the number of fixations and the average duration of these fixations. The introduced GRI is based on the relationship between both of these parameters and can be used as a single-value metric to assess visual expertise in vision-intensive domains, such as teaching. However, until now, the GRI has been a seldom-explored metric, and evidence is limited to lab-based on-action eye-tracking studies (Gegenfurtner et al., 2020; Grub et al., 2022). Therefore, the first aim of this project was to use the GRI to measure teachers' visual expertise based on real-world gaze data collected with a mobile eye-tracking device during instruction.

The second aim of the present study was to investigate the effect of various velocity thresholds for eye-movement identification on the eye-tracking parameter/GRI metric using the Identification by Velocity Threshold (I-VT) algorithm and fixation merging. This study is hypothesis driven and involves two related research questions:

1. Is the Gaze Relational Index (GRI) lower for experienced teachers as compared to novice teachers?

We aimed to explore the potential utility of GRI as an indicator of visual expertise. Based on previous findings, we hypothesized that experienced teachers use more top-down knowledge-based processing of visual information, leading to their ability to scan the visual field more rapidly. Thus, we expected more and faster fixations among experienced teachers. Novice teachers, in comparison, use more bottom-up salience-based processing of visual information, resulting in fewer and longer fixations. Therefore, we expect the GRI to be higher for novice teachers than for experienced teachers.

2. How do the eye-movement parameters (number of fixations and duration of fixations) and the GRI change…
   a) depending on the choice of velocity thresholds for eye movement detection based on the Identification by Velocity Threshold (I-VT) algorithm?
   b) depending on fixation merging?

We expected that the different velocity thresholds would lead to different results regarding the detection of fixations and saccades, thus affecting the GRI. Based on the logic behind the IV-T algorithm, we expected that the lower the selected velocity threshold, the fewer eye movements would be classified as fixations. However, based on prior eye-tracking protocols and studies (Andersson et al., 2017; Olsen, 2012), we hypothesize that this is not a linear process (i.e., the velocity threshold of 30°/s, as compared to 60°/s, does not classify half of the eye movements as fixations, mainly because, when using higher velocity thresholds, more smooth-pursuit eye movements and slow saccades were identified as fixations.

Furthermore, we expected that fixation merging would significantly reduce the number of fixations and, therefore, the GRI of a given participant. The extent to which outcomes differ is difficult to predict, so this research question is exploratory in nature.

Methods

Participants

The data were obtained from four in-service mathematics teachers (two females and two males). Each teacher gave a lesson ranging between 60 and 90 minutes in length in four different higher secondary schools (grade 9) in Germany. All participating teachers taught similar content (matrix calculus) during data collection. In addition, the sampled lesson was minimally predetermined to allow for some consistency across teachers and their individual lessons. Teachers were given 5 minutes of their class to recap the topic and tasks of the last lesson and the remaining time to introduce new content. Two of the participating teachers were novices, with average teaching experience of 1.55 years (SD = 0.05), while the other two teachers were experienced teachers, with average teaching experience of 11.21 years (SD = 0.8). The teachers were between 27 and 62 years old, (M = 37.25, SD = 16.64). Class sizes ranged from 14 to 24 students (M = 18.5, SD = 3.84). At the time of data collection, all teachers had known their students since the beginning of the school year (5 months).

Procedure

Mobile eye-tracking recording took place during a regular class period, which was chosen to interfere as little as possible with the regular lesson plan. We used a Tobii Pro Glasses 2 with a temporal resolution of 100 Hz to collect eye-movement data (Tobii, 2022). Before the recordings began, the eye-tracking glasses were calibrated until satisfactory calibration was achieved. All participating teachers were advised not to move their eye-tracking glasses during the recording of their eye movements. After the recording, the participating teachers were interviewed using a questionnaire (an assessment of the lesson, demographic data, and professional experience).

Data (Pre-)Processing

Data Collection

We exported the raw data using the Tobii Lab Analysis Software (Tobii, 2022), which gave us information about the eye and gaze positions at each recording timestamp and performed all subsequent fixation calculations in Python. For each timestamp, we recorded the time since the beginning of the recording in milliseconds, the pupil positions of the left and right eye at this timestamp in 3D space, the gaze points at this timestamp in 3D space, and a 2D representation of the gaze points at this timestamp. The duration of the recordings per participant varied between 24 and 68 minutes. To control for these time differences and limit their effect on the eye-tracking parameter, we extracted, for each person, all eye movements during the first 20 minutes of the recording and discarded the remainder of the data for our analysis.

Fixation-Classification Algorithm

Fixation calculation. We based our fixation calculation on the I-VT algorithm, as described by Salvucci and Goldberg (2000) and Olsen (2012). First, we calculated the point-to-point velocities for each pair of consecutive recording timestamps (t1, t2) by performing the following steps:

1. We calculated the timestamp of the exact point in time between t1 and t2 by taking the mean of t1 and t2.
2. We calculated the position of the left eye at the timestamp t1t2 by taking the mean of the left eye position vector at t1 and the left eye position vector at t2. We did the same for the right eye.
3. We calculated the visual angle between the left eye position at timestamp t1t2 and the gaze position at t1, as well as the gaze position at t2. We did the same for the right eye. This gave us an indicator of how far the gaze has moved from timestamp t1 to timestamp t2.
4. We divided the visual angle by the time elapsed between t1 and t2 in seconds. This gave us the angular velocity of an

eye movement in degrees/second at timestamp t1t2.

5. We aggregated the velocity of the left and right eye by taking the mean of both velocities. If the velocity of one eye could not be inferred (e.g., because the person had blinked with one eye at time t1 or t2 or both), we took the velocity of the other eye. If the velocity of neither eye could be inferred, the sample was declared an invalid value.
6. We calculated the gaze positions in 2D and 3D space at timestamp t1t2 by taking the mean of each gaze position at time t1 and time t2.
7. For each point, we stored the timestamp t1t2, the angular velocity v_t1t2 at this timestamp, the gaze points at t1t2, and the eye position at t1t2.
8. Next, we labeled all points with velocities below or equal to the velocity threshold parameter as fixations and all points with velocities above the threshold as a saccade. To study the effect of this velocity parameter on eye-tracking parameters, such as the duration and number of fixations, we performed our analysis with various threshold values between 10 and 150 (step size = 10). In addition, we paid special attention to the velocity threshold of 30°/sec, as this is used by default in the fixation filter of the Tobii Lab Analysis Software, and the velocity threshold of 100°/sec, as this is used by default in the attention filter of the Tobii Lab Analysis Software (Tobii, 2022).

Building fixation groups. After this, we built fixation and saccade groups by merging all consecutive points containing a fixation to a fixation group, all consecutive saccades to a saccade group, and all consecutive points with an invalid value to an invalid group. For each group, we defined the start time as the point in time between the timestamp of the first sample in this group and the timestamp of the last sample in the preceding group. Similarly, we defined the end time as the point in time between the timestamp of the last sample in this group and the timestamp of the first sample in the preceding group. We calculated the duration for the group by subtracting the start time from the end time, and we calculated the eye and gaze positions by taking the mean of all eye and gaze points in this group. We furthermore stored the eye-movement type (fixation, saccade, or invalid), as well as a counter for fixations, saccades, and invalid values.

Fixation merging. We then merged the fixation groups that were divided by a saccade or invalid value but were close in time and space. We did this via the following steps:

1. For each pair of subsequent fixation groups f1 and f2, we calculated the time between the end of f1 and the beginning of f2. If this time was shorter than a threshold (max_time_betw_fixations), we continued with Step 2; otherwise, we continued with the next fixation pair. We used a max_time_betw_fixations threshold of 75 milliseconds, as recommended in Olsen (2012).
2. We calculated the visual angle between f1 and f2 by using the mean eye position for f1 and f2, the gaze position for f1, and the gaze position for f2 for the left eye. We did the same for the right eye and merged the visual angles of both eyes as described in Step 5 of the fixation calculation. If the overall angle was shorter than a threshold value (max_angle_betw_fixations), we merged the fixation groups in the same way as consecutive fixations. All saccades and invalid values between f1 and f2 were thus discarded. We used a max_angle_bw_fixations threshold of 0.5 degrees, as recommended in Olsen (2012).

To study the effect of fixation merging on eye-tracking parameters, we performed our analysis once with and once without fixation merging.

Eye-tracking parameter calculation. We then continued to examine individual eye-tracking parameters. For each person, we calculated the number of fixations $fixnr_{Person\ x}$, as well as the mean fixation duration $fixdur_{Person\ x}$, meaning the sum of lengths of all fixations of this person divided by the number of fixations. Furthermore, we defined the GRI of a person as $GRI_{Person\ x} = fixnr_{Person\ x} / fixdur_{Person\ x}$ and calculated this index for each person.

To allow an expert-novice contrast, we calculated the mean fixation number, $fixnr_{Group\ x}$, by taking the mean of the fixation numbers of all participants in this group. Furthermore, we calculated the mean fixation duration of this group, $fixdur_{Group\ x}$, by taking the mean of the fixation durations of all participants in this group. We then calculated the GRI of a group, as defined by (Gegenfurtner et al., 2020), by using the following formula: $GRI_{Group\ x} = fixnr_{Group\ x} / fixdur_{Group\ x}$.

Results

Differences Between Experienced and Novice Teachers' GRI

The first research question examined the extent to which experienced and novice teachers differ in terms of the GRI. Table 1 shows the number of fixations, duration of fixations, and GRI using velocity thresholds of 30°/s and 100°/s and separated by expertise level. The descriptive results indicate that experienced teachers had more fixations, shorter fixation durations, and lower GRI values as compared to novice teachers. Although the trend observed in the results has not changed, varying velocity thresholds have an effect on the eye movement parameters/GRI. For example, while the difference between expertise groups in terms of GRI is marginal at a velocity threshold of 30°/s, novice teachers' GRI is more than double that of experienced teachers at a velocity threshold of 100°/s.

Table 1. Group-based eye-tracking parameter and GRI with velocity threshold of 30/100 and no merging of fixations

	Fixation Number		Mean Fixation Duration		GRI
	M	*SE*	*M*	*SE*	
Experts (VT 30°/s)	4319.50	566.50	125.52	17.82	0.030
Novices (VT 30°/s)	3991.50	569.50	156.35	3.35	0.039
Experts (VT100°/s)	3981.50	674.50	186.79	29.51	0.047
Novices (VT100°/s)	2569.50	401.50	326.37	18.70	0.127

Effect of Varying the Velocity Threshold on the Number of Fixations, Duration of Fixations, and GRI

The second research question examined the effect of varying the velocity threshold and fixation merging (yes/no) on eye movement parameters/GRI. Figure 1 shows the eye movement parameter/GRI for each person within each analysis. Merging the categories seems to have little to no effect on the mean fixation duration, fixation number, and GRI of a given person. However, as indicated in the results for RQ1, the velocity threshold seems to have a strong influence on eye movement parameters/GRI. Specifically, a higher velocity threshold leads to more samples being classified as fixation. Because consecutive samples containing a fixation are merged, a higher velocity threshold implies a higher mean fixation duration. At the same time, a higher velocity threshold leads to a lower number of fixations for thresholds above 30–40°/s. At first sight, this seems to be counterintuitive, but it is the case due to the merging of consecutive samples. For example, there could be three samples in the dataset, s1, s2, and s3. With a velocity threshold of 30°/s, these would be classified as s1 = fixation, s2 = saccade, and s3

= fixation, resulting in a fixation number of two. With a velocity threshold of 100°/s in contrast, s2 could be classified as fixation as well. As consecutive fixations are merged, this would result in a fixation number of one, meaning the fixation number decreases with an increasing velocity threshold. However, the above-identified relation is not linear, which means that the order of the participants in terms of their GRI values is changed. For example, E1 has a higher GRI than N2 when using a velocity threshold of 30°/s, but a lower GRI than N2 when using a velocity threshold of 100°/s.

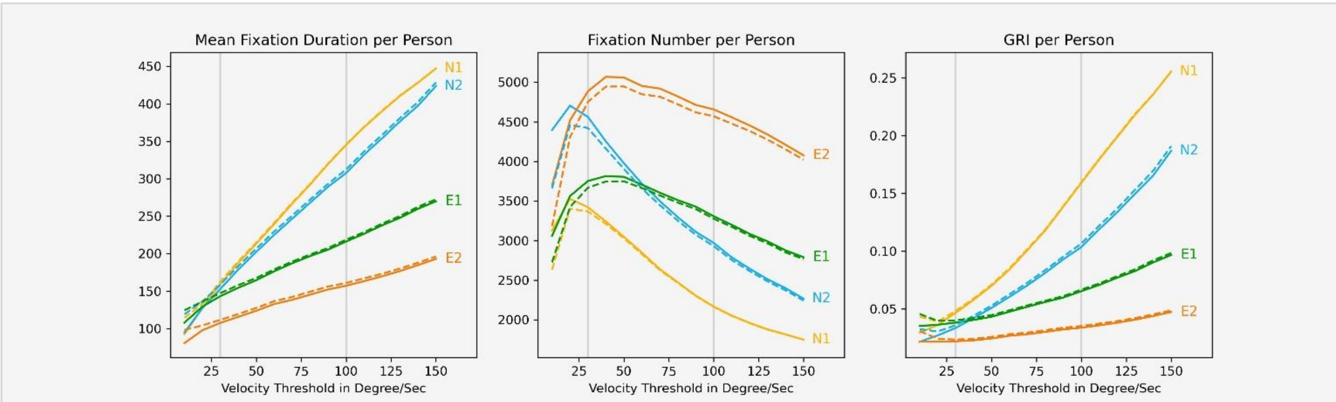

Figure 1. Mean fixation duration, fixation number, and GRI per person for a fixation calculation with different velocity thresholds between 10 and 150 °/s. Solid lines represent the analysis without the merging of fixation groups, and dashed lines represent the analysis with fixation-group merging. The velocity thresholds of 30°/s and 100°/s are marked with a grey line. E1/2 = experienced teachers; N1/2 = novice teacher.

Discussion

The teaching profession heavily depends on visual information. Teachers visually perceive, collect, and process information in a complex and dynamic classroom environment (Wolff et al., 2016). In recent years, cognitively oriented educational research found that experienced teachers develop domain-specific visual expertise that has not yet been developed in novice teachers (Kosel et al., 2021; van den Bogert et al., 2014; Wolff et al., 2016). The present study aimed to contrast the visual expertise of experienced and novice teachers, as measured by the GRI, in highly dynamic real-world classroom environments using mobile eye-tracking data. Furthermore, the study explores how different configurations (varying velocity thresholds and fixation-merging methods) of the IV-T algorithm for eye-movement classification affect the results of the study. In general, our findings reveal the perceptual superiority of domain experts, as indicated by lower GRI values, and suggest that the use of different velocity thresholds for eye-movement identification significantly affected the results of our study.

The GRI as an Expertise-sensitive Metric in Research About Teachers' Visual Expertise

We expected experienced teachers to process visual information more quickly and with more numerous fixations, indicating the domain-specific superiority of experienced teachers in terms of visual processing (Gegenfurtner et al., 2011) and, thus, require less time and effort to comprehend the complexity of classroom situations (Gegenfurtner et al., 2022). Therefore, the GRI (the ratio of the mean number of fixations to the mean duration of fixations) was expected to be lower for experienced teachers than for novices. We were able to provide support for this hypothesis, as we found that experienced teachers have more fixations and shorter average fixation durations than novice teachers and, thus, lower GRI values as compared to novice teachers. As compared to other studies in the context of visual expertise among medical experts and novices

(Gegenfurtner et al., 2020), as well as experienced and novice teachers (Grub et al., 2022), the calculated GRI values in this study were more sensitive to expertise.

One decisive reason for the heterogeneous results was that, as compared to the studies outlined above, we have begun to step outside of artificial classroom environments (laboratory setups) and move toward the more natural conditions teachers typically face in real classrooms using mobile eye-tracking. It has been shown that eye movements in the real world generally vary more significantly among participants (Dowiasch et al., 2020). Dowiasch and colleagues (2020) argue that this could be because mobile eye-tracking gaze recordings are generally much less restrictive than laboratory gaze recordings, allowing participants to behave more naturally. In this context, teachers often experience a much higher level of complexity in their real work environment, which is difficult to mirror in laboratory eye-tracking research. Therefore, the general transferability of results from eye-movement measurements taken in the laboratory to the real world seems difficult, although researchers must better understand visual behavior/expertise in natural environments (Dowiasch et al., 2020; Gegenfurtner et al., 2020). We have taken this step with this study and can confirm our assumptions about expertise differences based on the GRI. The results may indicate that experienced teachers' superior visual processing appears in complex and dynamic real-life situations.

Varying the Velocity Thresholds for Eye-movement Identification Influences the GRI

Across all research areas, eye-tracking-based studies face the critical challenge of transforming the raw gaze signals of the eye-tracker (i.e., gaze origin and gaze direction) into meaningful gaze parameters (i.e., fixations and saccades; Olsen, 2012; Tobii, 2022). There are numerous algorithms available for this task, and the algorithms often have various customizable configurations (Andersson et al., 2017). We investigated the effect of using different velocity-threshold settings on one of the most commonly used algorithms (IV-T; Andersson et al., 2017; Olsen, 2012) on the results of our mobile eye-tracking study (RQ2a). The results indicate that the choice of a velocity threshold influences the mean fixation duration and fixation number for a given person and, consequently, influences the GRI for a given person.

In addition, we found that the selection of the velocity threshold influences not only the absolute size of the GRI but also the rank order of participants regarding their GRI. In other words, the different velocity thresholds do not have a linear effect on the number and duration of fixations or GRI values. Concerning the default fixation filter (30°/s) and attention filter (100°/s) provided by Tobii (Olsen, 2012; Tobii, 2022), the results are less influenced in this regard when interpreted on the averaged-group level (experts versus novices) than when interpreted on the individual level (e.g., the comparison of individual participants). However, because eye-tracking studies have comparatively few participants as compared to other traditional study designs (i.e., questionnaire surveys), the presumed influence on the results is all the more striking, for example, when comparing group means.

Adopting this more methodological perspective, we argue that the heterogeneity that occurs in the results of visual expertise studies (e.g., as described by Klostermann & Moeinirad, 2020) regarding the number and duration of fixations among domain experts may be due not only to the different study contexts (e.g., varying professional domains or tasks) but also to the choice of a specific velocity threshold. Thus, the choice of velocity threshold should be regularly reported in publications. Furthermore, the process of fixation merging (RQ2b) did not affect our results. This is likely because, in our data, very few fixation groups are merged. Choosing higher parameters for the maximal time between fixations and the maximum angle between fixations would result in more fixations being merged and could, consequently, lead to larger differences between analyses. Because the manual settings of fixation merging are more restricted as compared to velocity thresholds in current software packages (Olsen, 2012; Tobii, 2022), we assume that the influence of fixation merging in studies is reduced because default values are often maintained.

In sum, our study demonstrates the importance of transparently specifying the configurations of algorithms for eye-movement classification in eye-tracking studies that base their interpretations on fixations and saccades. This is one step toward valid, reliable, and objective measurements of eye movements in the field of visual expertise. Based on our results and in agreement with Hossain and colleagues (2016), we recommend using the 100°/s fixation filter when mobile eye tracking is used and head movements are involved.

Limitations and Future Directions

The present study has four main limitations that can be addressed by future research. First, our study is limited to descriptive (group) comparisons, mainly due to its small sample size. Because the present results are exploratory, further research is needed to confirm these observed differences using a larger sample size.

Second, we have limited knowledge about the extent to which the GRI is related to the specific situations teachers face in the classroom. Therefore, the following considerations must be taken into account: our analysis showed that experienced and novice teachers differed in their visual behavior, as measured by the GRI, but we know little about *how* they differed in their interpretations of what they saw. Future research should focus on a more comprehensive combination of eye-tracking and think-aloud protocols to understand how the GRI relates to the underlying instructional situations that the teacher cognitively and visually faced during the eye-tracking recording. Another way to achieve this would be to code the first-person video recorded by the mobile eye-tracking afterward to investigate the GRI, for example, in different forms of instructional interaction (frontal instruction or group work). In this context, it should also be noted that we used a gaze-based approach (analyses are based only on the fixations) and have not integrated any areas of interest. This means that we did not consider the distribution of attention to specific areas in the classroom. This brings up an important point that should be considered in future research. To realize the full potential of the GRI, future studies should integrate relevant areas of interest in mobile eye-tracking data to analyze which areas in the classroom are being processed with a high or low GRI, for example, to analyze the GRI in relation to task-related and task-irrelevant areas. To summarize this above-described outlook, there is a further need to understand in which situations the visual expertise of experienced teachers comes to the fore.

Third, we have focused on two essential parameters (velocity threshold and fixation merging) for the various configurations of the IV-T algorithm and ignored other aspects, such as interpolation (filling gaps in raw eye-tracking data in which no signal was recorded) or active noise reduction (e.g., noise may be caused by imperfect system settings; Tobii, 2022). Remaining to be clarified is the extent to which these, often manufacturer-specific, methods of data preprocessing influence the results, especially when studies use different eye-tracking devices.

Fourth, due to the relatively small sample size in our exploratory study, we were unable to statistically control for potentially confounding variables, such as teacher experience or class size. Although the descriptive analysis indicates small standard deviation values for teacher experience and class size, suggesting a minimal influence in this study, it is essential for future replication studies to address these confounding variables using regression analysis and larger samples with greater variance.

Finally, GRI values depend on the length of the recording. While the mean fixation duration should remain relatively constant over a period of time, the number of fixations increases with each recording minute, leading to a decrease in GRI values. Therefore, we recommend analyzing participants over the same amount of time and stating the recording time when reporting a study.

Conclusion

Our study has two important consequences concerning research on teachers' visual expertise. First, the GRI may serve as a sensitive measure of visual expertise when using mobile eye-tracking data. The lower GRI of experienced teachers indicates that they have distinct visual behavior, which is indicative of their fine-grained domain-specific knowledge organization and reflected in their visual expertise. Regardless of the velocity thresholds chosen for identifying fixations, the experienced teachers showed shorter and more fixations, which resulted in lower GRI values. However, from a methodological viewpoint, the study also showed that the selected velocity threshold parameters alter and, in the worst-case scenario, bias the results of an eye-tracking study. Therefore, in the interest of the further generalizability of the results within visual expertise research, researchers are encouraged to be transparent about reporting their algorithm configurations regarding eye-movement identification.

Authors' Declarations

The authors declare that there are no personal or financial conflicts of interest regarding the research in this article.

The authors declare that the research reported in this article was conducted in accordance with the Ethical Principles of the *Journal of Expertise*.

The authors declare that they are not able to make the dataset publicly available but are able to provide it upon request.

ORCID iDs

Christian Kosel
https://orcid.org/0000-0002-4857-4148

Tina Seidel
https://orcid.org/0000-0002-2578-1208

Jürgen Pfeffer
https://orcid.org/0000-0002-1677-150X

References

Aglioti, S., Cesari, P., Romani, M., & Urgesi, C. (2008). Action anticipation and motor resonance in elite basketball players. *Nature Neuroscience*, *11*, 1109–1116. https://doi.org/10.1038/nn.2182

Andersson, R., Larsson, L., Holmqvist, K., Stridh, M., & Nyström, M. (2017). One algorithm to rule them all? An evaluation and discussion of ten eye movement event-detection algorithms. *Behavior Research Methods*, *49*(2), 616–637. https://doi.org/10.3758/s13428-016-0738-9

Berliner, D. C. (2001). Learning about and learning from expert teachers. *International Journal of Educational Research*, *35*(5), 463–482. https://doi.org/10.1016/S0883-0355(02)00004-6

Chaudhuri, S., Muhonen, H., Pakarinen, E., & Lerkkanen, M.-K. (2021). Teachers' focus of attention in first-grade classrooms: Exploring teachers experiencing less and more stress using mobile eye-tracking. *Scandinavian Journal of Educational Research*, *9891*(2), 1–17. https://doi.org/10.1080/00313831.2021.1958374

Chi, M. T. H., Glaser, R., & Farr, M. J. (Eds.). (1988). The nature of expertise. Lawrence Erlbaum Associates, Inc.

Cortina, K. S., Miller, K. F., McKenzie, R., & Epstein, A. (2015). Where low and high inference data converge: Validation of CLASS assessment of mathematics instruction using mobile eye tracking with expert and novice teachers. *International Journal of Science and Mathematics Education*, *13*(2), 389–403. https://doi.org/10.1007/s10763-014-9610-5

Dolezalova, J., & Popelka, S. (2016). ScanGraph: A novel Scanpath comparison method using visualisation of graph cliques. *Journal of Eye Movement Research*, *9*(5), 1–13. https://doi.org/10.16910/jemr.9.4.5

Dowiasch, S., Wolf, P., & Bremmer, F. (2020). Quantitative comparison of a mobile and a stationary video-based eye-tracker. *Behavior Research Methods*, *52*(2), 667–680. https://doi.org/10.3758/s13428-019-01267-5

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, *102*(2), 211–245. https://doi.org/10.1037/0033-295X.102.2.211

Gegenfurtner, A. (2020). *Professional vision and visual expertise* (Habilitation Thesis). University of Regensburg, Regensburg.

Gegenfurtner, A., Boucheix, J.-M., Gruber, H., Hauser, F., Lehtinen, E., & Lowe, R. (2020). The Gaze Relational Index as a measure of visual expertise, *Journal of Expertise*, *3*, 32–40.

Gegenfurtner, A., Gruber, H., Holzberger, D., Keskin, Ö., Lehtinen, E., Seidel, T., Stürmer, K., & Säljö, R. (2023). Towards a cognitive theory of visual expertise: Methods of inquiry. In C. Damşa, A. Rajala, G. Ritella, & J. Brouwer (Eds.), *Re-theorising learning and research methods in learning research* (pp. 146–163). Routledge. https://doi.org/10.4324/9781003205838-10

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

Gegenfurtner, A., & Seppänen, M. (2013). Transfer of expertise: An eye tracking and think aloud study using dynamic medical visualizations. *Computers & Education*, *63*, 393–403. https://doi.org/10.1016/j.compedu.2012.12.021

Goodwin, C. (1994). Professional vision. *American Anthropologist*, *96*(3), 606–633. https://doi.org/10.1525/aa.1994.96.3.02a00100

Grub, A.-S., Biermann, A. & Brünken, R. (2020). Process-based measurement of professional vision of (prospective) teachers in the field of classroom management. A systematic review. *Journal for Educational Research Online*, 12(3), 75–102. https://doi.org/10.25656/01:21187

Grub, A.-S.., Biermann, A., Lewalter, D., & Brünken, R. (2022). Professional vision and the compensatory effect of a minimal instructional intervention: A quasi-experimental eye-tracking study with novice and expert teachers. *Frontiers in Education*, *7*, 1,248. https://doi.org/10.3389/feduc.2022.890690

Holmqvist, K., Nystrom, M., Andersson, R., Dewhurst, R., Jarodzka, H., & van de Weijer, J. (2015). *Eye tracking: A comprehensive guide to methods and measures*. Oxford: Oxford University Press.

Hossain, A., & Miléus, E. (2016). Eye movement event detection for wearable eye trackers. Thesis, Applied Mathematics, MAI, Linköpings Universitet.

Huang, Y., Miller, K. F., Cortina, K. S., & Richter, D. (2021). Teachers' professional vision in action. *Zeitschrift für Pädagogische Psychologie*, *36*, 1–18. https://doi.org/10.1024/1010-0652/a000313

Klostermann, A., & Moeinirad, S. (2020). Fewer fixations of longer duration? Expert gaze behavior revisited. *German Journal of Exercise and Sport Research*, *50*(1), 146–161. https://doi.org/10.1007/s12662-019-00616-y

Kosel, C., Holzberger, D. & Seidel, T. (2021). Identifying expert and novice visual scanpath patterns and their relationship to assessing Learning-Relevant student characteristics. *Frontiers in Education*, *5*. https://doi.org/10.3389/feduc.2020.612175

Kosel, C., Böheim, R., Schnitzler, K., Holzberger, D., Pfeffer, J., Bannert, M. & Seidel, T. (2023). Keeping track in classroom discourse: comparing in-service and pre-service teachers' visual attention to students' hand-raising behavior. *Teaching and Teacher Education, 128*, 104142. https://doi.org/10.1016/j.tate.2023.104142

Lachner, A., Jarodzka, H., & Nückles, M. (2016). What makes an expert teacher? Investigating teachers' professional vision and discourse abilities. *Instructional Science*, *44*(3), 197–203. https://doi.org/10.1007/s11251-016-9376-y

Lappi, O., Rinkkala, P., & Pekkanen, J. (2017). Systematic observation of an expert driver's gaze strategy—an on-road case study. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.00620

Lowe, R. K., & Boucheix, J.-M. (2016). Principled animation design improves comprehension of complex dynamics. *Learning and Instruction, 45,* 72-84. https://doi.org/10.1016/j.learninstruc.2016.06.005

McIntyre, N. A., & Foulsham, T. (2018). Scanpath analysis of expertise and culture in teacher gaze in real-world classrooms. *Instructional Science*, *46*(3), 435–455. https://doi.org/10.1007/s11251-017-9445-x

Olsen, A. (2012). The Tobii I-VT Fixation Filter [Computer software]: Tobii.

Panetta, K., Wan, Q., Rajeev, S., Kaszowska, A., Gardony, A. L., Naranjo, K., Taylor, H., Agaian, S. (2020). ISeeColor: Method for advanced visual analytics of eye tracking data. *IEEE Access*, *8*, 52278–52287. https://doi.org/10.1109/ACCESS.2020.2980901

Peißl, S., Wickens, C. D., & Baruah, R. (2019). Eye-tracking measures in aviation: A selective literature review. *The International Journal of Aerospace Psychology*, *28*(3–4), 98–112. https://doi.org/10.1080/24721840.2018.1514978

Salvucci, D. D., & Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols. In A. T. Duchowski (Ed.), *Proceedings of the symposium on eye tracking research & applications - ETRA '00* (pp. 71–78). New York, New York, USA: ACM Press. https://doi.org/10.1145/355017.355028

Seidel, T. & Stürmer, K. (2014). Modeling and measuring the structure of professional vision in preservice teachers. *American Educational Research Journal, 51*(4), 739–771. https://doi.org/10.3102/0002831214531321

Seidel, T., Schnitzler, K., Kosel, C., Stürmer, K., & Holzberger, D. (2021). Student characteristics in the eyes of teachers: Differences between novice and expert teachers in judgment accuracy, observed behavioral cues, and gaze. *Educational Psychology Review*, *33*(1), 69–89. https://doi.org/10.1007/s10648-020-09532-2

Tobii (2022). Tobii Pro Lab User Manual [Computer software]: Tobii.

Van den Bogert, N., van Bruggen, J., Kostons, D., & Jochems, W. (2014). First steps into understanding teachers' visual perception of classroom events. *Teaching and Teacher Education*, *37*(1), 208–216. https://doi.org/10.1016/j.tate.2013.09.001

Van Es, E. A., & Sherin, M. G. (2010). The influence of video clubs on teachers' thinking and practice. *Journal of Mathematics Teacher Education*, *13*(2), 155–176. https://doi.org/10.1007/s10857-009-9130-3

Weibel, N., Fouse, A., Emmenegger, C., Kimmich, S., & Hutchins, E. (2012). Let's look at the cockpit. In C. H. Morimoto, H. Istance, J. B. Mulligan, P. Qvarfordt, & S. N. Spencer (Eds.), *Proceedings of the Symposium on Eye Tracking Research and Applications - ETRA '12* (p. 107). New York, USA: ACM Press. https://doi.org/10.1145/2168556.2168573

Wolff, C. E., Jarodzka, H., van den Bogert, N., & Boshuizen, H. P. A. (2016). Teacher vision: Expert and novice teachers' perception of problematic classroom management scenes. *Instructional Science*, *44*(3), 243–265. https://doi.org/10.1007/s11251-016-9367-z

Received: 18 August 2022
Revision received: 19 July 2023
Accepted: 19 July 2023