Using eye-tracking to investigate teacher gaze: Data accuracy and drawing of meaningful dynamic areas of interest in video stimuli

Christina Ioanna Pappa, Christian Kosel, Katharina Schnitzler, Tina Seidel

Technical University Munich, Germany

Abstract: Studies that investigate teacher attention via eye-tracking methodology display variety in their reporting styles and consideration of their data quality. This may be due to the fact that eye-tracking has been newly introduced in teacher research, and systematic guidelines are not yet established. This especially accounts for the influence of the quality of the raw data (i.e., accuracy level) and the way that the raw data is processed through the drawing of dynamic areas of interest (AOIs) in video stimuli. The present study investigates the influence of various accepted accuracy levels on the number of fixations and three variations of AOI drawings (student shape, indicated by outlined areas; face, indicated by ovals; and student area, indicated by rectangles) on common eye-tracking metrics: number of fixations, glances, and fixation duration. Sixty-two participants observed a video stimulus with five marked students as the targeted AOIs. A one-way ANOVA was conducted to examine the influence of different accuracy levels ($>1°$, $>0.5°$ to $\leq 1.0°$, and $\leq 0.5°$) of the data on the number of fixations, while the effect of three different dynamic AOI shapes (student shape, face, rectangle) was investigated with a series of repeated-measure one-way ANOVAs. The results indicated no significant difference between the accuracy levels and the number of fixations. For the different AOI shapes, significant differences were observed. When using rectangles, more fixations and glances were recorded in contrast to the other two forms. The average fixation duration was greatest when only the faces were marked. This indicates that depending on the research question and the position of the AOIs, researchers may choose different forms of AOIs and consider the accuracy of their data.

Eye-tracking technology is a valuable tool to examine learning and attentional processes in educational science (Jarodzka et al., 2017; Lai et al., 2013; Shayan et al., 2017). Although eye-tracking has been used in cognitive

studies for many decades, it is relatively new in educational research regarding teacher attentional processes (e.g.,Cortina et al., 2015; Van den Bogert et al., 2014; Yamamoto & Imai-Matsumura, 2013). Previous studies used teachers' eye movements to investigate attentional processes regarding classroom management, cultural differences and differences among novice and expert teachers (e.g.,Cortina et al., 2015; McIntyre & Foulsham, 2018; Stürmer et al., 2017).

In educational science, many eye-tracking studies regarding human cognition and perception used static stimuli such as images (Rayner et al., 2007; Reichle, 2006) and texts (Parkhurst & Niebur, 2003). In contrast, studies regarding teachers' attentional processes used either dynamic video stimuli (e.g.,Van den Bogert et al., 2014; Wolff et al., 2016; Yamamoto & Imai-Matsumura, 2013) or real classroom situations (e.g., Cortina et al., 2015; McIntyre & Foulsham, 2018; McIntyre et al., 2019; Stürmer et al., 2017). For these types of stimuli, some methodological issues remain about the quality of the raw data and the standardization of the analysis (Holmqvist et al., 2011; Holmqvist et al., 2012), because there are so far no specific guidelines and systematic investigations. As a result, these past studies demonstrate a great variability when reporting the accuracy of the data and the position and shape of the Areas of Interest (AOIs), a frequent method used to analyze the eye-tracking data.

Therefore, the present study aimed to investigate two important parameters regarding the quality of a study, the accuracy of the data (i.e., quality of the raw data) and the shape of the AOIs (i.e., quality of the data extraction). More specifically, the study investigated the effect of accuracy levels >1°, > 0.5° to ≤ 1.0°, and ≤ 0.5° on the number of fixations and the influence of the AOI shape on the number of fixations, glances, and average fixation duration.

This study aims to provide understanding regarding the significance of the quality of raw data as well as the quality of extraction of the data. It may serve as an explorative study on how to treat and analyze eye-tracking data in applied research designs. Moreover, results may function as an evidence-based guide for future educational research regarding the drawing of meaningful AOIs in video stimuli.

1 Theoretical background

1.1 Eye-tracking as a measure of attention

In recent years, eye-tracking technology has been used more frequently in educational research not only to examine learning processes (Jarodzka et al., 2017; Lai et al., 2013) but also as a method to investigate the teacher attention. Eye movements give insight into the cognitive processes and allow conclusions about the subject's attention, as the eye-mind hypothesis suggests (Cortina et al., 2015; Just & Carpenter, 1976). Specifically, the time that the eye remains fixated on a subject reflects information about people's cognitive and attentional processes toward an object (Just & Carpenter, 1976; Rayner, 2009). Fixations are considered as a visual gaze on a steady location during which information is noticed and cognitively processed (Holmqvist et al., 2011). This linkage between eye fixations and attention is evident especially in complex tasks such as reading and visual search (Rayner, 2009; Rayner & Castelhano, 2008).

To investigate attentional processes, research focused commonly on the investigation of fixations with a focus on the number of fixations and duration of fixations (Stürmer et al., 2017; Van den Bogert et al., 2014; Wolff et al., 2016; Yamamoto & Imai-Matsumura, 2013). The number of fixations (i.e., fixation count) refers to the absolute number of fixations on a specific object. The time that the eye is fixated on a certain location indicates how long information is being perceived and cognitively processed (Rayner, 2009). Therefore, both duration of fixations and number of fixations can be used as a measure of one's attention toward a specific object or area (Holmqvist et al., 2011).

A common way to analyze the number of fixations in eye-tracking research is to use the so-called Areas of Interest (AOIs). The AOIs can be defined as the segmenting regions that the researcher is interested in, which allow to analyze and simplify the data in an illustrative way such as graphs, strings, and transition matrices (Holmqvist et al., 2011). Therefore, researchers can identify how often a participant fixated a specific object displayed in an AOI.

1.2 Eye-tracking quality and validity of results

Data quality is a necessary parameter to establish the validity of eye-tracking results. There are several aspects which could affect the quality of the data,

such as the properties of the eye-tracking apparatus, the environmental conditions of the location of the experiment (e.g., lighting), the position of the devices and also the calibration process (Holmqvist et al., 2011; Holmqvist et al., 2012; Nyström et al., 2013; Ooms et al., 2015). Calibration process refers to the procedure in which the subjects have to look at some predefined targets, and for each target, the eye-tracker associates the subject's eye image position with the location of the target (Nyström et al., 2013).

Inaccuracy and imprecision of the calibration can affect the quality of the research results. Precision can be defined as the consistency of the calculated gaze points when the true gaze direction is stable (Holmqvist et al., 2012, p. 46). During the eye-tracking, tracking ratio, the amount of gaze positions that the eye tracker recorded during the entire experiment (Amso et al., 2014; Holmqvist et al., 2011; SMI, 2014), influences data quality. Besides these issues of the eye-tracking process, also the extraction of eye movements through AOIs may influence the quality of the eye-tracking data used in subsequent analyses.

1.3   Accuracy: The need for standardized accuracy quality criteria

Data accuracy, also known as offset, is considered an essential property of data quality in eye-tracking research. Data accuracy can be defined as the distance in degrees of visual angle between the actual gaze target and the recorded position (horizontal x-axis and vertical y-axis) (Holmqvist et al., 2011; Holmqvist et al., 2012). High accuracy in the eye-tracking data is necessary to create accurate scientific assertion (Nyström et al., 2013). Although manufacturers state that the average level of accuracy of their systems is better than 0.5°, prior research using images and videos with dynamic objects (e.g., cars, aeroplanes) as stimuli reveals that the accuracy levels are in reality around 1°, after the invalid recordings are excluded (Komogortsev & Khan, 2008; Nyström et al., 2013; Zhang & Hornof, 2011).

Lack of accuracy might be especially problematic in stimuli in which the AOIs are close to each other, such as in reading, or unevenly distributed, such as on a web page. Small ambiguity such as 0.5° to 1° might be critical for correct eye-tracking data analysis (Blignaut & Wium, 2014). In situations where the AOIs are positioned close to each other, fixations, which belong to one AOI, could be attributed to another AOI due to their local proximity, and as a result, the systematic error could cause misinterpretations (Zhang & Hornof, 2011).

Even though the calibration procedure is often used to minimize any systematic error, the eye-trackers frequently sustain some degree of misinterpretation (Blignaut & Wium, 2014). Therefore, it is important to report the accuracy of the raw data used for analyses because reporting only the manufacturer's information on accuracy can be misleading, since accuracy can differ among the participants and the experimental conditions (Blignaut & Beelders, 2009).

1.4 Areas of interest (AOIs): The challenge of drawing AOIs

Both static as well as dynamic AOIs (i.e., segmenting regions that the researcher is interested in) are currently used to analyze teacher eye-tracking data (e.g., Stürmer et al., 2017; Van den Bogert et al., 2014; Wolff et al., 2016; Yamamoto & Imai-Matsumura, 2013). Static stimuli (e.g., texts) are often defined through scripts, which assign fixations to AOIs based on their position on the visual plane, in contrast to dynamic stimuli (e.g., videos) in which AOIs are usually drawn and defined by hand (Orquin et al., 2016). In video stimuli, AOIs have to be adapted manually to the pictured subjects' movements. In situations when the AOIs are adjusted automatically, they still require subsequent manual checks.

Prior research reported that there is a lack of common methlology for defining the AOIs, because the AOI method has been frequently recreated by software developers and researchers (Holmqvist et al., 2011). As a result, there is a variety of how AOIs are drawn and defined across studies (Holmqvist et al., 2011; Orquin et al., 2016). Holmqvist and colleagues (2011) recommended, in regard to the size of the AOIs, that they should be drawn with a buffer space of 1° to 1.5 ° visual angle around the subject. Moreover, in case of low accuracy, the margin space can be increased in order to include all the fixations that belong to the subject. However, Orquin and colleagues (2016) noted, this suggestion might be critical in situations where the subjects are close to each other and especially in video stimuli, where dynamic AOIs are drawn around people. By increasing the AOI margins, false fixations might be ascribed to an AOI, to which they do not belong.

Although, Orquin and colleagues (2016) conducted a systematic investigation of the AOI size and margin for static stimuli such as texts and pictures, eye-tracking research remains unclear about the AOI size and shape in video stimuli. Results of their study on static stimuli suggested keeping

a larger AOI margin in situations where the distance between the AOIs is large. In contrast, when the AOIs are close to each other, and the fixations are likely to overlap, smaller margins should be applied in order to balance the range of true and false fixations. Additionally, results from a survey regarding the use of AOIs in behavioral eye-tracking studies showed a lack of general agreement among researchers on the use of AOIs (Orquin et al., 2016).

1.5 Measuring teacher attention

Although eye-tracking technology has been used in empirical studies about human cognition and perception for many decades (Shayan et al., 2017), it is relatively new to educational research, particularly, regarding teacher attention. Eye-tracking is considered a low-inferential way to investigate teacher attention regarding classroom management (Stürmer et al., 2017), cultural differences (McIntyre & Foulsham, 2018; McIntyre et al., 2019; McIntyre et al., 2017), and variations among novice and expert teachers (Cortina et al., 2015; Van den Bogert et al., 2014; Wolff et al., 2016). Within the field, the eye-tracking method has been used with dynamic video stimuli, and both with remote eye trackers (Van den Bogert et al., 2014; Wolff et al., 2016; Yamamoto & Imai-Matsumura, 2013) as well as during authentic classroom situations with mobile eye-tracking glasses (Cortina et al., 2015; McIntyre & Foulsham, 2018; McIntyre et al., 2019; Stürmer et al., 2017).

In these previous studies, eye-tracking data were mainly analyzed using the indicators of the number of fixations, duration of fixations, and uniformity of the distribution of fixations (e.g.,Stürmer et al., 2017; Van den Bogert et al., 2014; Wolff et al., 2016; Yamamoto & Imai-Matsumura, 2013). However, because of the lack of information and no specific evidence in the literature about the accuracy of the data and the AOI size and shape in video stimuli, past studies demonstrate a great variability regarding the reported accuracy of the data and the position and the shape of the AOIs.

Regarding data accuracy, one group of researchers reported the accuracy level of their apparatus to be horizontal x-axis 0.56° and vertical y-axis 0.40° (McIntyre & Foulsham, 2018; McIntyre et al., 2019; McIntyre et al., 2017). Moreover, previous studies found various ways of how to construct meaningful AOIs from polygonal forms in which several students are included to AOIs adjusted to single students' shape. Unfortunately, the decision processes behind the construction were not provided in those publications,

making it difficult to compare the different advantages and disadvantages of each individual way of AOI construction. AOIs were mostly reported in a figure, without specific information about their size, position, and implemented adjustments for the video dynamics such as students moving around or raising their hands. However, some reported having large enough and spatially distinguishable AOIs (Stürmer et al., 2017; Van den Bogert et al., 2014; Wolff et al., 2016; Yamamoto & Imai-Matsumura, 2013).

In general, eye-tracking technology could allow more accurate measurement of cognitive processes and conclusions about subject's attention, but some methodological issues should be considered regarding the standardization of the analysis and the quality of the data (Holmqvist et al., 2011; Holmqvist et al., 2012). The field, therefore, might profit from setting joint quality criteria to be able to obtain comparable quality indicators and to replicate studies (Holmqvist et al., 2012).

1.6   The current study

The lack of a common standardization of AOI shapes leads to problems concerning comparability and interpretation of results across studies. This study aimed to explore how to treat and evaluate the eye-tracking data in an applied setting. When evaluating the eye-tracking data, several decisions have to be made regarding the data quality, such as whether the accuracy is great enough and whether all the participants should be included in the data analysis. Additionally, the researcher has to decide how to evaluate the data efficiently, i.e., in a short time but with high quality, as well as which AOIs should be drawn and used, since manually drawing the dynamic AOIs could be a time-consuming procedure. Thereby, the current study addresses the following two questions with high practical relevance:

RQ1: Do the different accuracy levels (>1°, > 0.5° to ≤ 1.0°, and ≤ 0.5°) influence the number of identified fixations of dynamic AOIs?

Although to date there are no previous publications that investigated this issue for dynamic AOIs, it is expected that the greater the accepted accuracy level (horizontal x-axis and vertical y-axis) is, the lesser the number of fixations (i.e., fixations inside an AOI) that are recorded (H1). This could be attributed to the deviation of the recorded fixations, and more specific, the participant might have looked at an AOI, but because of the low accuracy,

the fixation was not recorded inside the AOI. Therefore, a larger accepted accuracy level, could lead to fewer fixations recorded.

RQ2: Does the use of different shapes (i.e., AOI in a form of rectangle, AOI individually drawn around the students' figure, and AOI in the form of an oval drawn around student's face) of dynamic AOIs in classroom video stimuli influence eye movement metrics, such as a) number of fixations, b) number of glances, and c) average fixation duration?

It is expected that large AOI shapes of rectangles record the greatest number of fixations (H2a) and glances (H2b), compared to smaller AOIs (i.e., students' figure, face AOIs). The bigger AOIs might include other objects as well, such as parts of other students who are sitting closely, or working materials lying on the table. Thereby, the recorded fixations and glances on these other objects might be counted to the respective student AOI, which they do not belong to (Orquin et al., 2016; Zhang & Hornof, 2011), (i.e false positives). While a smaller AOI such as students face may miss some fixations (i.e false neagtives).

Regarding the average fixation duration, it is assumed that longest fixations might be recorded on the AOI focusing on students' faces (H2c). As past research revealed, faces attract people's attention (Borji et al., 2013; Cerf et al., 2009; Judd et al., 2009). Moreover, research from Gullberg and Holmqvist (2001) noted that people spend more time fixating faces than gestures. Despite the fact that the other AOI shapes include students' gesture, gestures do not last for a long time. Therefore, it is expected that the participants might spend more time fixating on students faces.

2   Methods

2.1   Participants

The initial sample consisted of 77 participants (M_{Age} = 28.14; SD_{Age} = 8.47; 65.8% female, 32.9% male); due to low tracking ratio (≤ .90) (i.e., the amount of gaze positions that the eye tracker recorded during the entire experiment) 15 participants were excluded (Holmqvist et al., 2011). Therefrom the sample included for analyses consisted of 62 participants (M_{Age} = 28.50; SD_{Age} = 8.99). The sample is composed of 34 pre-service teachers who participated in the study as a part of one of their courses, as well as 11 in-service teachers, and

17 researchers in the field of teacher education who were recruited via their public email addresses and participated voluntarily in the study.

2.2   Apparatus

Stimuli were presented using Experiment Center 3.7. Eye movements were recorded with a SMI RED 500 binocular remote Eye-Tracker with a 22-inch display monitor and a sampling frequency of 500 Hz. The participants were seated 65 cm in front of the Eye-Tacker. Ceiling light with closed blinds was used to maintain a stable light condition. In order to increase precision, a height-adjustable table was adjusted for each participant. A chin rest was used to limit strong head movements (Nyström et al., 2013). Before presenting the video, a 9-point system-controlled calibration was performed, followed by a 4-point validation.

2.3   Procedure

Data were collected from May to December 2018 with each participant seen individually in the laboratory. Participants watched a trailer video (2:30 min.) to get familiar with the classroom environment and the lesson topic. The eye-tracking was conducted while the subjects watched a stimulus video (11 min.). The video displayed an 8th-grade mathematics lesson about the geometry topic of figure scaling from a German high school, with five marked students. The lesson was videotaped from the right corner of the classroom that consisted of 23 students (9 female and 14 male). Participants were instructed to observe these five students attentively in order to be able to solve a diagnostic task afterward. The five marked students were distributed in the classroom so that they represented challenges that occur when implementing eye-tracking in real classroom situations in which students may be sitting close, behind each other, and sometimes overlapping and hiding each other. This realistic position of the marked students is chosen in order to investigate the situations where the AOIs are close to each other, and therefore, false fixations could be attributed to other AOIs, and as a result, the systematic error could cause misinterpretations (Orquin et al., 2016; Zhang & Hornof, 2011).

2.4 Eye-tracking metrics: Number of fixations, glances, and average fixation duration

The analysis of the current study was based on the first three minutes of the video because this part showed an introduction to the topic followed by a lively classroom discussion where the students raised their hands and participated verbally. This situation was chosen because it is especially challenging to draw adjusted dynamic AOIs for situations in which students move their bodies. To analyze the eye-tracking data the software BeGaze was used. The five marked students were used as targets for the AOIs (see Figure1-3). The exactness of the AOIs were manually adjusted for every single video frame.

Fixation count (i.e. fixations inside specific AOIs), glances count (i.e. how often an AOI was visited), and the average fixation durations (i.e. the average length of the fixations within an AOI) were calculated with the BeGaze system defaults. In case that the marked students (AOIs) overlapped each other (see Figure 3), the AOI metrics were counted towards the AOI in the foreground only (SMI, 2014). Because the current study focused on the AOI metrics of all marked students in total, it was important that the metrics towards the AOIs were counted only once. For example, in a situation when a student raised his hand and covered his classmate, the fixations were counted for the student that raised his hand because he was in front of the other student. In the rectangle AOIs (see Figure 3), the rectangle 4 was selected to be in the foreground because the student in this AOI is less visible in comparison to the student in rectangle 3 AOI.

2.5 Data analysis

As a preliminary analysis, we investigated whether participants varied in their vision patterns due to their different professions. Thereby, a one-way analysis of variance (ANOVA) was used to examine differences in the number of fixations between pre-service teachers, in-service teachers, and teacher educators. The homogeneity of variance has been tested and, because of large sample size differences, Hochberg's GT2 posthoc test was conducted. The results reveal that the assumption of homogeneity of variance was violated; therefore, the Brown-Forsythe F-ratio is reported. Because of the fact that there are no statistically significant differences between the different professions of the participants (i.e., pre-service teachers, in-service teachers, and teacher educators) (Brown-Forsythe -F(2, 26.19) = 01.53 p = .23), the sample was not divided in the further data analyses.

The influence of data accuracy on the number of fixations

To answer the first research question three accuracy levels (i.e., >1°, > 0.5° to ≤ 1.0°, and ≤ 0.5°) were examined. The first accuracy level was set to 0.5°, meaning that deviation on horizontal x-axis and vertical y-axis from the target point during validation should be lower or equal to 0.5°. This value was chosen because most manufacturers report that their systems have an average accuracy level better than 0.5°. Second accuracy level was set to include all participants with a deviation above 0.5° and a maximum of 1° on horizontal x-axis and vertical y-axis. The second accuracy level group was chosen because previous research revealed that often the accuracy levels are around 1°, after invalid recordings are eliminated (Komogortsev & Khan, 2008; Nyström et al., 2013; Zhang & Hornof, 2011). The third accuracy level was 1°, including the participants with a deviation greater than 1° to investigate the difference on the results when the accuracy of that data is not controlled. In the current study, the participants reached a maximum deviation of 1.5°.

To examine the influence of the accuracy levels, AOIs of student's figure shape were used (see Figure 1). Previous research has shown that both faces as well as gestures are attractors of peoples' attention (Gullberg & Holmqvist, 2001). Therefore, the student figure AOI shape was chosen for the accuracy analysis because it is assumed that this shape contains not only students' facial expressions but also gestures, and therefore it was considered the AOI shape that might represent the most of the relevant information.

For the analysis, the three different accuracy levels were examined. Within the sample 18 participants ($M_{Fixation\ Count}$ = 303.88; $SD_{Fixation\ Count}$ = 63.11) had an accuracy level of ≤ 0.5°, 38 participants ($M_{Fixation\ Count}$ = 301.02; $SD_{Fixation\ Count}$ = 76.06) reached an accuracy level between > 0.5° and ≤ 1.0°, and 6 participants an accuracy > 1° ($M_{Fixation\ Count}$ = 239.90; $SD_{Fixation\ Count}$ = 136.26).

One-way analysis of variance (ANOVA) was used to examine the influence of the different accuracy levels on the number of fixations. The homogeneity of variance has been tested, and Hochberg's GT2 posthoc tests were conducted (because of large sample size differences), to test for significant differences between the three accuracy levels (i.e., >1°, > 0.5° to ≤ 1.0°, and ≤ 0.5°).

The influence of the shape of dynamic AOIs on the number of fixations, glances and the average fixation duration

To answer the second research question, regarding the influence of the shape of dynamic AOIs on the eye movement metrics, three different AOI shapes were drawn. For the first scenario, AOIs were drawn manually in the form of the student's shape (Figure 1), in order to keep lower AOI margins, but included both students' faces and gestures. Second, only the faces of the five students were marked as AOIs in an oval shape (Figure 2) since prior research revealed that faces attract people's attention (Borji et al., 2013; Cerf et al., 2009; Judd et al., 2009). And for the third scenario, rectangles were used as AOI forms (Figure 3) because it is a quick method to draw AOIs that includes students' faces, and also their bodies. The oval and rectangle AOIs (Figure 2, 3) were automatic AOI shape forms in BeGaze software and therefore their drawing and shape adjustments in each video frame was a faster and easier process in comparison to the manually drawn AOI shape in the form of the student's shape (Figure 1).

Figure 1. First AOI scenario: AOIs indicated by grey outline highlighting students (including shape and face).

Figure 2. Second AOI scenario: AOIs indicated by grey ovals highlighting students' faces.

Figure 3. Third AOI scenario: AOIs indicated by grey rectangles highlighting students.

A series of repeated-measures one-way ANOVAs were used to examine the differences in the AOI shape on the number of fixations, number of glances, and fixation duration. The normal distribution has been tested for the different eye-tracking metrics of the three AOI shapes, which were used as the repeated within-subject factor in the ANOVA. Afterward, Bonferroni post hoc analyses were conducted to investigate which metrics among the three AOI shapes differ from each other in detail.

3   Results

3.1   The influence of various accuracy levels on the number of fixations

Regarding the first research question about the influence of the accuracy levels on the number of fixations of the student's figure AOI shape, the results reveal that the assumption of homogeneity of variance was violated; therefore, the Brown-Forsythe F-ratio is reported. Contrary to our expectations, no statistically significant differences between the three accuracy levels >1°, > 0.5° to ≤ 1.0°, and ≤ 0.5° were found (Brown-Forsythe -$F(2, 8.38) = 0.97$, $p = .42$) (see Figure 4).
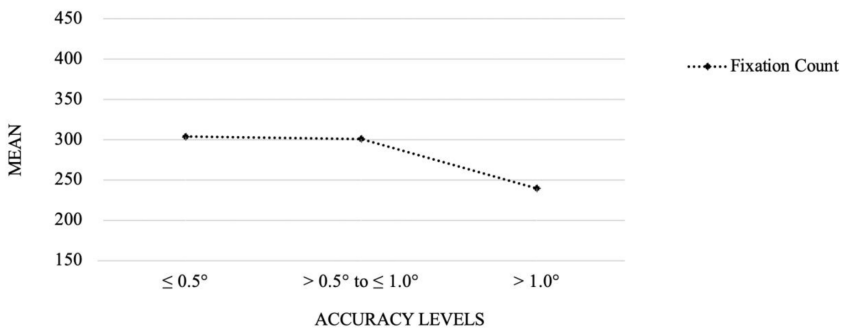
Figure 4. Application of the three different accuracy levels (i.e., >1°, > 0.5° to ≤ 1.0°, and ≤ 0.5°) and the differences in recorded number of fixations.

3.2   The influence of the shape of dynamic AOIs on the number of fixations, glances and the average fixation duration

Differences in number of fixations, number of glances, and average fixation duration across the three AOI scenarios are displayed in Figure 5.

Number of fixations

The three different shapes of the AOIs showed a significant main effect on the number of fixations (*Greenhouse-Geisser-F*(1.40, 84.40) = 383.19, $p < .001$, $\eta P^2 = .86$). A post hoc Bonferroni analysis indicated that significantly ($p < .001$) more fixations were recorded in the student's figure AOI scenario (M = 296.06, *SD* = 10.40) than in the face AOI scenario (*M* = 221.32, *SD* = 12.14). In addition, significantly ($p < .001$) more fixations were recorded in the rectangle AOI scenario (*M* = 422.08, *SD* = 14.30) than student's figure and face.

Number of glances

Regarding the number of glances, the repeated-measures one-way ANOVA revealed significant differences in the number of glances according to different shapes of AOIs (*Greenhouse-Geisser-F* (1.28, 77.10) = 113.03, $p < .001$, $\eta P^2 = .65$). A post hoc Bonferroni analysis indicated that significantly (p < .001) more glances were recorded in the student's figure AOI scenario (*M* = 177.49, *SD* = 5.88) than in the face AOI scenario (*M* = 151.31, *SD* = 7.13), and significantly ($p < .001$) more number of glances recorded in the rectangle AOI scenario (*M* = 208.57, *SD* = 6.77) than student's figure and face.

Average fixation duration

Repeated-measures one-way ANOVA results show a significant difference in the average fixation duration according to different shapes of AOIs *Greenhouse-Geisser-F*(1.25, 75.08) = 23.80, $p < .00$, $\eta P^2 = .28$.

A post hoc Bonferroni analysis indicated that the duration of fixations in the face AOI scenario were significantly longer (*M* = 392.34, *SD* = 18.13, $p < .001$) than at the student's figure AOI scenario (*M* = 365.89, *SD* = 14.91, $p < .001$) and the rectangle AOI scenario (*M* = 352.81, *SD* = 13.98, $p < .000$).
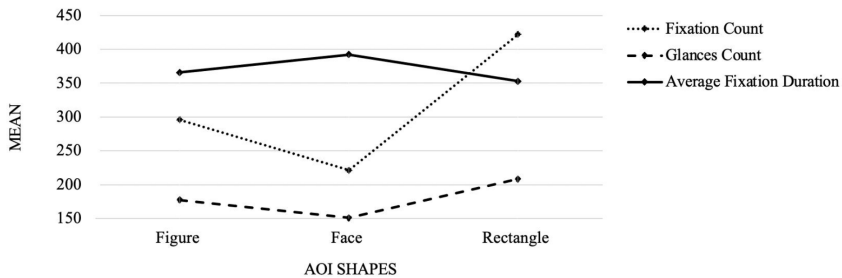
Figure 5. Application of three different AOI shapes (figure, face, rectangle) and differences in recorded eye movement indicators (fixations, glances, average fixation duration).

4   Discussion

As has been mentioned before, there are methodological issues concerning the quality of teacher eye movement raw data and the standardization of the analysis (Holmqvist et al., 2011; Holmqvist et al., 2012) since, so far, there have been no specific guidelines and systematic investigations. Thereby, the current study aimed to investigate two important parameters of the quality of a study, the accuracy of teacher eye movement data (i.e., quality of the raw data) and the used shapes of AOIs (i.e., quality of the data extraction). The current study investigated the influence of various accuracy levels on the number of fixations and of the AOI shape on the number of fixations, glances, and the average fixation duration. The study adds the following findings to the research field: First, the results revealed that three accuracy levels (i.e., $>1°$, $> 0.5°$ to $\leq 1.0°$, and $\leq 0.5°$) do not result in systematic effects on the number of recorded fixations. Second, regarding the influence of the shape of dynamic AOIs on the eye-tracking metrics, the results indicate that the way AOI shapes are used might have an effect on the number of recorded fixations, glances and the average fixation duration.

4.1   Central findings

Regarding the accuracy of the data, the results are inconsistent with the first Hypothesis (H1), as there was no significant difference between the accuracy levels in the number of fixations. Since there is no previous research that

investigated this issue on dynamic AOIs, that question aimed to explore how to analyze the eye-tracking data from an applied study and whether the accuracy levels of the raw data could influence the eye movement metrics. Therefore, based on the non-significant results about the influence of the accuracy level on the number of fixations using the AOI shape on students' figure, it was decided to use the whole sample for further analysis and include all participants. Moreover, based on the non-significant results, it could be assumed that further studies could use the eye-tracking data from an applied study without applying further accuracy quality criteria and excluding participants with higher accuracy deviation. However, although the accuracy level was not controlled, the participants of this study reached a maximum deviation of 1.5°. Thereby, further simulation studies should investigate whether there is a threshold of the accuracy level where differences might become significant.

In order to be able to set consistent standards about the accuracy of data and compare the results across studies, reporting only the accuracy level of the eye-tracking system might not be enough since accuracy may vary among the participants and the experimental conditions (Blignaut & Beelders, 2009). Therefore, based on previous research (Blignaut & Wium, 2014; Holmqvist et al., 2011; Nyström et al., 2013; Ooms et al., 2015) and the current findings on dynamic AOIs, it could be suggested for further studies to report the accuracy level of their data after the invalid recordings (e.g., data with low tracking ratio) are excluded. Nevertheless, to set standard quality criteria, further research should investigate other important parameters regarding the eye-tracking data quality, such as the precision of the data and their influence on the eye movement metrics on dynamic AOIs, to enhance guidance on what quality criteria to report and consider when analyzing eye-tracking data. For example, further studies should consider developing a systematic guide with step-by-step procedures in order to enhance the guidance on what quality criteria to report and how to calculate, for example, the precision of the data. Moreover, another interesting aspect for future research regarding the data accuracy would be to manipulate the raw data by adding different offsets to investigate its influence on eye movement metrics.

The second finding of this study refers to the influence of the shape of the AOIs on the number of recorded fixations and glances. Consistent with the hypothesis 2a and 2b for the AOIs in a rectangle form most fixations and

glances were recorded. Therefore, it could be assumed that in situations where the dynamic AOIs are close to each other, the AOI with the larger shape might record the most number of fixations and glances. In contrast, the AOI with the smallest shape (i.e., the AOI on students' faces) could record fewer number of fixations and glances. However, as previous studies noted, in situations with larger AOIs placed close to each other, false fixations could be attributed to an AOI, which they do not belong to. Hence, in situations with video stimuli where students are used as dynamic AOIs, false fixations and glances might be ascribed to wrong students and consequently, the results could lead to inaccurate assumptions (Orquin et al., 2016; Zhang & Hornof, 2011). Based on the results above, it is suggested that for the number of fixations and glances analysis, the AOI shape on students' figure might be more suitable in situations where people are used as targeted AOIs. The AOI shape on students' figures contains information about gestures and facial expressions. Moreover, the AOI shape on students' figure has smaller margins, and as previous research suggested, when the AOIs are close to each other and the fixations are likely to overlap, smaller margins should be used, to balance the range of true and false fixations (Orquin et al., 2016). Thereby, by using the AOI shape on student's figure, researchers could avoid misleading assumptions due to false assigned fixations and glances.

The third finding refers to the influence of the AOI shape on the average fixation duration. Confirming hypothesis 2c, the longest duration was recorded on the AOI drawn on students' face. As mentioned before, previous research revealed that faces attract people's attention (Borji et al., 2013; Cerf et al., 2009; Judd et al., 2009) in comparison to gestures (Gullberg & Holmqvist, 2001). Therefore, the amount of time that the participants fixating at the student's face was longer than the rest of the body because hand gestures do not last for a long time. Hence, it is suggested for further studies to consider carefully what information they are interested in and to take into account that the metrics of fixation duration is influenced by what parts of students are included in AOIs.

Future research could continue the systematic investigation of the influence of the shape of dynamic AOIs in classroom video stimuli to develop a standardized way of drawing dynamic AOIs. This would diminish problems with the comparability of results across studies. Further research could complement our applied findings, for example, by an investigation of effects

of AOI sizes on eye movements with simulation studies where AOIs are normalized and effects of different sizes are compared. Moreover, individual studies investigating eye-tracking data should carefully decide which AOI shapes they use and justify their choice to increase transparency and comparability across studies.

4.2 Limitations

Some limitations have to be considered when interpreting the results. First, findings of the current applied study should be considered as preliminary insights for the field of teaching and they should be further investigated by systematic simulation studies as outlined previously. Furthermore, results related to the data quality are limited to the data accuracy. Although we are aware that precision is an important additional quality criterion for eye-tracking data, it was not included in the present study due to the fact that the level of precision must be calculated manually and is not automatically provided within common eye-tracking or analysis software. As reports on studies concerned with teacher eye-movements mostly lack information even for the level of accuracy of their eye-tracking data, we argue that an awareness for this indicator and report of this information in publications could be seen as a first step to advance the quality of studies within this field. Third, findings relate to the AOIs refer only to the shape of the AOIs. The study did not investigate the size of the AOIs.

4.3 Conclusion

The present study investigated two important parameters of the quality of a study, the accuracy of the data and the shape of the AOIs. It provided evidence that there are no significant differences between various accepted accuracy levels on number of fixations. Nevertheless, it is suggested for further research to report data accuracy after the invalid recordings are excluded in order to gain more systematic knowledge about variations in data accuracy. Furthermore, the study provides evidence that the AOI shapes influence the number of fixations, glances, and the average fixation duration. Therefore, decisions about AOI shapes should be made transparent in publications, and possible limitations depending on chosen AOI shapes should be discussed.

References

Amso, D., Haas, S., & Markant, J. (2014). An eye tracking investigation of developmental change in bottom-up attention orienting to faces in cluttered natural scenes. *PloS one*, *9*(1), e85701. https://doi.org/https://doi.org/10.1371/journal.pone.0085701

Blignaut, P., & Beelders, T. (2009). The effect of fixational eye movements on fixation identification with a dispersion-based fixation detection algorithm. *Journal of Eye Movement Research*, *2*(5). https://doi.org/https://doi.org/10.16910/jemr.2.5.4

Blignaut, P., & Wium, D. (2014). Eye-tracking data quality as affected by ethnicity and experimental design. *Behavior research methods*, *46*(1), 67–80. https://doi.org/https://doi.org/10.3758/s13428-013-0343-0

Borji, A., Sihite, D. N., & Itti, L. (2013). What stands out in a scene? A study of human explicit saliency judgment. *Vision research*, *91*, 62–77. https://doi.org/https://doi.org/10.1016/j.visres.2013.07.016

Cerf, M., Frady, E. P., & Koch, C. (2009). Faces and text attract gaze independent of the task: Experimental data and computer model. *Journal of vision*, *9*(12), 10–10. https://doi.org/https://doi.org/10.1167/9.12.10

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/https://doi.org/10.1007/s10763-014-9610-5

Gullberg, M., & Holmqvist, K. (2001, April). Visual attention towards gestures in face-to-face interaction vs. on screen. In *International Gesture Workshop*, 206–214. https://doi.org/https://doi.org/10.1007/3-540-47873-6_23

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures.* OUP Oxford. http://lup.lub.lu.se/record/1852359

Holmqvist, K., Nyström, M., & Mulvey, F. (2012). Eye tracker data quality: What it is and how to measure it. *Proceedings of the symposium on eye tracking research and applications,* 45–52. https://doi.org/https://doi.org/10.1145/2168556.2168563

Jarodzka, H., Gruber, H., & Holmqvist, K. (2017). Eye tracking in educational science: Theoretical frameworks and research agendas. *Journal of Eye Movement Research*, *10*(1). https://doi.org/http://dx.doi.org/10.16910/jemr.10.1.3

Judd, T., Ehinger, K., Durand, F., & Torralba, A. (2009). *Learning to predict where humans look.* In IEEE 12th International Conference on Computer Vision, Kyoto, Japan. https://doi.org/10.1109/ICCV.2009.5459462

Just, M. A., & Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cognitive psychology*, *8*(4), 441–480. https://doi.org/https://doi.org/10.1016/0010-0285(76)90015-3

Komogortsev, O. V., & Khan, J. I. (2008). Eye movement prediction by Kalman filter with integrated linear horizontal oculomotor plant mechanical model. In *Proceedings of the 2008 symposium on Eye tracking research & applications* (pp. 229–236). https://doi.org/https://doi.org/10.1145/1344471.1344525

Lai, M.-L., Tsai, M.-J., Yang, F.-Y., Hsu, C.-Y., Liu, T.-C., Lee, S. W.-Y., Lee, M.-H., Chiou, G.-L., Liang, J.-C., & Tsai, C.-C. (2013). A review of using eye-tracking technology in exploring learning from 2000 to 2012. *Educational research review*, *10*, 90–115. https://doi.org/https://doi.org/10.1016/j.edurev.2013.10.001

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/https://doi.org/10.1007/s11251-017-9445-x

McIntyre, N. A., Jarodzka, H., & Klassen, R. M. (2019). Capturing teacher priorities: Using real-world eye-tracking to investigate expert teacher priorities across two cultures. *Learning and Instruction*, *60*, 215–224. https://doi.org/https://doi.org/10.1016/j.learninstruc.2017.12.003

McIntyre, N. A., Mainhard, M. T., & Klassen, R. M. (2017). Are you looking to teach? Cultural, temporal and dynamic insights into expert teacher gaze. *Learning and Instruction*, *49*, 41–53. https://doi.org/https://doi.org/10.1016/j.learninstruc.2016.12.005

Nyström, M., Andersson, R., Holmqvist, K., & Van De Weijer, J. (2013). The influence of calibration method and eye physiology on eyetracking data quality. *Behavior research methods*, *45*(1), 272–288. https://doi.org/https://doi.org/10.3758/s13428-012-0247-4

Ooms, K., Dupont, L., Lapon, L., & Popelka, S. (2015). Accuracy and precision of fixation locations recorded with the low-cost Eye Tribe tracker in different experimental setups. *Journal of eye movement research*, *8*(1). https://doi.org/10.16910/jemr.8.1.5

Orquin, J. L., Ashby, N. J., & Clarke, A. D. (2016). Areas of interest as a signal detection problem in behavioral eye-tracking research. *Journal of Behavioral Decision Making*, *29*(2–3), 103–115. https://doi.org/https://doi.org/10.1002/bdm.1867

Parkhurst, D., & Niebur, E. (2003). Scene content selected by active vision. *Spatial vision*, *16*(2), 125–154. https://doi.org/https://doi.org/10.1163/15685680360511645

Rayner, K. (2009). The 35th Sir Frederick Bartlett Lecture: Eye movements and attention in reading, scene perception, and visual search. *Quarterly journal of experimental psychology*, *62*(8), 1457–1506. https://doi.org/https://doi.org/10.1080/17470210902816461

Rayner, K., & Castelhano, M. S. (2008). Eye movements during reading, scene perception, visual search, and while looking at print advertisements. In M. Wedel & R. Pieters (Eds.), *Visual marketing: From attention to action* (pp. 9–42). Taylor & Francis Group/Lawrence Erlbaum Associates. https://psycnet.apa.org/record/2007-15678-002

Rayner, K., Li, X., Williams, C. C., Cave, K. R., & Well, A. D. (2007). Eye movements during information processing tasks: Individual differences and cultural effects. *Vision research*, *47*(21), 2714–2726. https://doi.org/https://doi.org/10.1016/j.visres.2007.05.007

Reichle, E. D. (2006). Computational models of eye-movement control during reading: Theories of the" eye-mind" link. *Cognitive systems research*, *7*(1), 2–3. https://doi.org/https://doi.org/10.1016/j.cogsys.2005.07.001

Shayan, S., Abrahamson, D., Bakker, A., Duijzer, C. A., & van der Schaaf, M. (2017). Eye-tracking the emergence of attentional anchors in a mathematics learning tablet activity. In *Eye-tracking technology applications in educational research* (pp. 166–194). IGI Global. https://doi.org/10.4018/978-1-5225-1005-5.ch009

SMI, S. I. (2014). *BeGaze Manual:* Version 3.4. Retrieved 20.10.2020 from https://psychologie.unibas.ch/fileadmin/user_upload/psychologie/Forschung/N-Lab/SMI_BeGaze_Manual.pdf

Stürmer, K., Seidel, T., Mueller, K., Häusler, J., & Cortina, K. S. (2017). What is in the eye of preservice teachers while instructing? An eye-tracking study about attention processes in different teaching situations. *Zeitschrift für Erziehungswissenschaft*, *20*(1), 75–92. https://doi.org/https://doi.org/10.1007/s11618-017-0731-9

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*, 208–216. https://doi.org/https://doi.org/10.1016/j.tate.2013.09.001

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

Yamamoto, T., & Imai-Matsumura, K. (2013). Teachers' gaze and awareness of students' behavior: Using an eye tracker. *Comprehensive Psychology*, *2*, article 6. https://doi.org/https://doi.org/10.2466/01.IT.2.6

Zhang, Y., & Hornof, A. J. (2011). Mode-of-disparities error correction of eye-tracking data. *Behavior research methods*, *43*(3), 834–842. https://doi.org/https://doi.org/10.3758/s13428-011-0073-0

Authors

Christina Ioanna Pappa, Technische Universität München, TUM School of Education, Professur für Technikdidaktik, Arcisstraße 21, 80333 München, Germany, e-mail: christ.pappa@tum.de

Christian Kosel, Technical University of Munich, Friedl Schöller Endowed Chair for Educational Psychology, Arcisstraße 21, 80333 München, Germany, e-mail: christian.kosel@tum.de

Katharina Schnitzler, echnical University of Munich, Friedl Schöller Endowed Chair for Educational Psychology, Arcisstraße 21, 80333 München, Germany, e-mail: katharina.schnitzler@tum.de

Tina Seidel, echnical University of Munich, Friedl Schöller Endowed Chair for Educational Psychology, Arcisstraße 21, 80333 München, Germany, e-mail: tina.seidel@tum.de

Využití eye-trackingu pro výzkum pohledu učitelů: Přesnost dat a vykreslení smysluplných dynamických oblastí zájmu v rámci video stimulů

Abstrakt: Studie, které zkoumají pozornost učitelů skrze eye-tracking vykazují velkou různorodost co se týče stylu reportování a pohledu na kvalitu dat. Důvodem může být to, že eye-tracking je v oblasti výzkumu učitele poměrně novou metodou a systematická metodologická doporučení prozatím neexistují. To se týká především kvality sbíraných dat (tedy přesnosti) a způsobu jejich zpracování skrze vykreslení dynamických oblastí zájmu (areas of interest, AOI) ve video stimulech. Předkládaná studie zkoumá vliv různých hladin akceptované přesnosti dat na počet fixací a tři způsoby vykreslení AOI (AOI ve tvaru žáka kreslená pomocí křivek; AOI tváře žáka kreslená pomocí oválů; AOI oblasti žáka vykreslená obdélníkem) na standardní eye-trackingové proměnné: počet fixací, počet pohledů (glances) a délka fixace. 62 účastníků sledovalo video stimul s pěti označenými žáky, kteří představovali cílové

AOI. Byla provedena jednoduchá analýza rozptylu s cílem určit vliv různých hladin přesnosti dat (>1°, > 0.5° to ≤ 1.0°, and ≤ 0.5°) na počet ficaxí. Vliv tří typů vykreslení dynamických AOI byl zkoumán pomocí několika jendofaktorových analýz rozptylu s opakovaným měřením. Výsledky neukázaly významný rozdíl mezi jednotlivými hladinami přesnosti. Významné rozdíly byly ale pozorovány u různých typů AOI. Při použití obdélníků bylo zaznamenáno více fixací a více návštěv AOI než u dalších dvou typů AOI. Nejdelší průměrné fixace byly zaznamenány u AOI zaměřených pouze na tváře žáků. Tyto výsledky nazančují, že je nutné zvážit výběr tvaru AOI v závislosti na výzkumné otázce a pozice AOI v prostoru a zároveň věnovat pozornost přesnosti dat.

Klíčová slova: eye-tracking, pozornost učitelů, přesnost dat, AOI, dynamický video stimul